

OM of: US-09-303-518d-653 to: A\_Geneseq\_032802.\* out\_format : pfs  
Date: Jun 30, 2002 7:14 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=firemap.n2p.spool -DEV=x11  
-O=/cgn2.1/USPNC\_model/US09303518/runat\_28062002.142712.4291/app.query.fasta.1.23501  
-DB=A\_Geneseq\_032802 -OFMT=fasta -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FCAPOP=6.000 -FCGAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500  
-DELLOC=6.000 -DELLOC=7.000 -START=1 -MATRIX=blsuum62  
-TRANS=human40.cdi -LIST=100 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pts  
-NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09303518 -CGCN=1.0 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-653  
Query Length: 4407  
Database: A\_Geneseq\_032802.\*  
Database sequences: 747574  
Database length: 111073796  
Search time (sec): 627.340000

Score list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:AA07304		1614.00	2049.86	3.2e-107	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA75364		1585.50	2013.94	3.4e-105	14
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA55292		1052.50	1327.66	6.1e-67	13
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA552745		1048.00	1321.84	1.3e-66	13
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:AA08135		721.50	916.78	2.0e-23	7
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA08135		442.50	548.38	2.9e-23	34
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA08135		355.00	432.61	3.8e-17	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081830		310.50	372.82	1.0e-13	12
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081828		310.50	372.77	1.0e-13	12
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081828		308.50	369.26	1.5e-13	13
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081609		306.00	366.89	2.2e-13	12
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081609		306.00	364.12	2.4e-13	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081609		296.50	352.36	1.1e-12	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081609		287.00	340.49	5.4e-12	14
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081609		286.50	339.85	5.9e-12	14
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081825		283.50	338.04	9.0e-12	12
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081825		283.00	337.34	9.8e-12	12
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081848		282.50	334.69	1.1e-11	14
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081848		279.00	335.41	1.7e-11	91
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081848		275.50	329.54	2.2e-11	14
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081848		275.50	330.77	3.0e-11	92
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081848		274.00	328.97	3.9e-11	91
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081827		273.00	323.10	5.1e-11	14
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081827		273.00	322.94	4.7e-11	94
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081827		272.00	325.34	5.6e-11	10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081840		272.00	325.27	5.6e-11	10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081840		271.50	314.12	9.4e-11	25
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081837		269.00	321.40	8.6e-11	11
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081837		269.00	321.34	8.7e-11	14
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AA081835		268.50	316.23	1.2e-10	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081846		268.50	316.23	1.2e-10	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AA081846		267.50	300.58	2.3e-10	57
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AA081846		267.00	314.30	1.5e-10	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081845		266.00	315.86	1.6e-10	11
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081845		266.00	315.75	1.6e-10	11
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081725		265.00	313.01	1.8e-10	15
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081725		265.00	308.44	2.1e-10	23
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081835		265.50	316.02	1.7e-10	10

/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081834	265.50	315.97	1.7e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:AA081723	265.00	311.72	2.1e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081723	261.00	303.26	4.6e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081708	261.00	298.77	5.4e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA0817403	260.50	290.60	7.6e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA0820365	258.00	306.33	5.8e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081833	257.50	306.64	6.1e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081832	257.50	306.58	6.1e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081830	257.50	297.17	8.1e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081847	256.50	304.45	7.5e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081843	256.00	304.85	7.5e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081842	256.00	304.79	7.8e-10
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081836	254.00	294.44	1.4e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:AA081833	253.50	292.22	1.6e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081826	253.00	295.72	1.5e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081826	251.50	294.97	1.9e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081826	251.50	294.97	1.9e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081826	248.00	297.20	2.7e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081835	247.50	286.03	4.2e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081849	245.00	288.92	5.8e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081849	243.50	283.50	7.4e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AA081838	243.50	281.28	8.0e-09
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	237.50	271.11	2.4e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081839	237.00	280.76	1.8e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA081838	237.00	280.69	1.8e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	237.00	278.32	1.9e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	236.50	270.32	2.7e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	236.50	270.32	2.7e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	235.00	274.66	4.2e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	233.00	270.01	4.2e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AA081838	230.00	261.64	8.1e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	230.00	260.93	8.3e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	229.00	260.93	8.4e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	228.50	265.75	8.4e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	228.50	261.88	9.6e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	228.50	261.88	9.6e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	228.50	261.58	9.7e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	228.50	261.19	9.9e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	227.50	264.47	9.9e-08
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	224.00	265.03	1.5e-07
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA081838	224.00	258.39	1.9e-07
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:AA081838	224.00	249.54	2.6e-07
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AA081838	223.00	260.33	1.8e-07
/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AA081838	223.00	254.54	2.4e-07

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:AA07304  
seq\_documentation\_block:  
ID: AA07304 standard; protein: 1541 AA.

AA07304:  
31-JAN-1991 (first entry)  
Igal protease.  
Igal: vaccine; meningitis; gonorrhoea; allergies.  
Haemophilus influenzae.  
W09011367-A.  
04-OCT-1990.  
16-MAR-1990: 90W0-DK00073.  
17-MAR-1989: 89DK-0001308.

PA (KILL/) KILIAN M.  
 XX Kilian M, Poulsen K;  
 PI  
 XX  
 DR WPI: 1990-320267/42.  
 DR N-PSDB: AAO06164.  
 XX  
 PT Immunoglobulin A1 protease prodn. - by cloning from  
 PT microorganisms for immunisation against immunoglobulin A1  
 PT protease producing bacteria  
 XX  
 PS Disclosure, fig 3, 44pp; English.  
 XX  
 CC This immunoglobulin (Ig)A1 protease is produced by recombinant DNA  
 CC methods. It is useful in a vaccine for e.g. meningococcal meningitis,  
 CC gonorrhea or allergic diseases. It specifically cleaves the heavy  
 CC chain of human IgA1 in the hinge region.  
 XX  
 SQ Sequence 1541 AA;

alignment\_scores:  
 Quality: 1614.00 Length: 1716  
 Ratio: 1.732 Gaps: 48  
 Percent Similarity: 54.312 Percent Identity: 26.865

alignment block:  
 US-09-303-518D-653 x AAR07304 ..

Align seg 1/1 to: AAR07304 from: 1 to: 1541

```

64 CGCTTCGCGCGCGCTTACCATATGCGTGTGCGCATTTGCC 113
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
5  LysPheLysLeuAsnPhelLeuThrValAlaTyralaLeuThr 21
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
114 CCAAGCCCGGCGGACACACTTATTCGCGCATACCTACCAATGATC 163
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
21  OTYrThrGluAlaAlaLeuValaIArgAspAspValaAspTyrGlnLeuPhe 38
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
164 GCGACTTTCGCGGAAATATAAGCAAGTTTCGCGCGGCGCAATATT 213
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
38  rGAspPheAlaGluAsnLysGlyLysPheSerValaGlyAlaThrAsnVal 54
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
214 GAGGTTTACAAACAAAAGGAGGTGTCGCGCAATCGATGACGAAAGC 263
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
55  LeuValLysAspLysAsnAsnLysAspLeuGlyThraAlaLeuProAsnG 71
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
264 C...CCGATGATGATTTTCTGTGTATCG...CGTACGCGGTGGCGG 307
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
71  yLePheMetLeaSprPheSerValaIAspValaAspLysArgIleAlaT 88
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
308 CATTCGCGCGGATCAATATATGTCGCGCGACAT...AACGGGCGC 354
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
88  hrLeuIleAsnProGlnTyrValaValaGlyValaLysHisValSerAsnGly 104
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
355 TATAACAATGTTGATTTTGGT.....GCGGAGGAGACAA 389
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
390 TCCGATGACGACGCG.....TTTCTTACCAAAATG 421
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
121 nAlaLysAlaHisArgAspValaSerSerGluGluAsnArgTyrPheSerV 138
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
422 TGAAGAAGAAATATATTAAGCAGGACTAACGCG..... 456
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
138 alGluLysAsnGlnTyrProThrLysLeuAsnGlyLysThrValThrThr 154
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
457 .....CATCCTTATGCGCGGATATATCATATATGCGCGTTTGA 494
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
495 CAATTTTGTACAGATGCAAGAACCTGTGAG...ATGACCGATTATATAG 541
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

```

171 pLysPheValThrGluValaIaProIleGluAlaSerThrAlaSerSerA 188
542 ATGGGTGGAATATCCCTGATTTAATTAATACCGATCGTTCGATC 591
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValaArgLeu 204
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
592 GAGCAGCGCAGACAAATATGCGGCTGTGATGAGACGAA..... 630
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
205 GlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuI 221
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
631 .CCCAATAACCGCGAAAGTTTCATATCAT.....ATGCAACGC 667
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
221 LeuAsnAsnHisGluValaGlyGlyAsnAsnLeuLysLeuValaGlyAspA 238
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
668 CATATTCCTGGCTCGCGTGGCAATACCTTTCACAAATGATGATCAGT 717
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
238 IarTyrThrTyrGlyIleAlaGlyThrProTyrLysValaHisHisGluAsn 254
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
718 GGTGCGACAGTCAACTAGTAGGCAAAAATTAACATPAGC...CCATA 764
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGlnHisSerAspProLy 271
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
765 TGGTTTTTACCAACAGAA.....GGCTCATTTGGCGACA 799
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValaLeuGlyAspS 288
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
800 GTGCGTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTAAAT 849
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
288 erGlySerProLeuPheValaTyrAspArgGluLysGlyLysTyrLeuPhe 304
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
850 AATGGGCTATTGCAACAGCAAGCAACCCCTTATAGCAAAAAGCAATGGCT 899
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
305 LeuGlySerTyrAspPheThrAlaGlyTyrAsnLysLysSer..... 318
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
900 CCAAGTAGTTCGTAAGATGCG...TTCATGATGAAATCTTGCTGGAG 946
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
319 .....TTPGlnGlnTyrAsnIleTyrLysSerGlnPheThrLysA 332
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
947 ATACCCATTACAGTATTCAGCAACCATCAAAATGCGAAATATCTTTT 996
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
332 spValLeu..... 334
997 AACGCAATATATATGCGCGCAGCAAAAATGCAATGCCAATATACACTA 1046
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
335 .....AsnLysAspSerAlaGlySerLeuIleGlySerLysThrAspTy 349
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1047 TTCTTACTTATAGATTAATAAACAGAACGCT.....CAAT 1084
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
349 rSerTyrPheSerAsnGlyLysThrSerThrIleThrGlyGlyLys 366
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAACCTGTTATGATGCT 1134
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
366 erLeuAsnValaAspLeuAlaAspGlyLysAspLysPro..... 378
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1135 GCAGTGGGGTCAACAGTTATGACCCAGACCTGAATATGAGAAATAT 1184
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
379 .....AsnHisGlyLysSerVa 384
1185 TTCCCTTATGCAAAAGAAAGTGAATGATTAATACACCAACATCA 1234
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
384 lThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsnAla 399
1235 ACCAAGCGCGGCGGCTTTGTATTTGAGGATATTTTACGCTC...TGC 1281
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
399 spGlnTyrAlaGlyLeuPhePheGlnGlyAspTyrGlnValaGlyGly 415
1282 CTTAAACACAGCAAGCGTGGCAAGCGGCGGCTCATATGATGATGATG 1331
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
416 ThrSerAspAsnThrThrTyrLysGlyAlaGlyValaSerValaIaGlu 432
1332 CAGTACCGCTTACTTGAAGTAACGCGTGGCAAGACCGCTGTCCA 1381
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
432 yLysThrValThrTyrLysValaHisAsnProGlnTyrAspArgLeuAla 449

```

1382 AATTCGCAAGGACGCTGTGTTCAAGCCAAAGGGGAAACCAAGGC 1431  
|||||  
449 ysllelylsglythrlleuilevalgluglythrglyaspasnlysgly 465  
1432 TGGGTACGGGTGGGAGCGGTAAAGTCATTGATCAGACGCGGAGCA 1481  
|||||  
466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGly 482  
1482 TCAAGGCAAAAACAAGCCTTAGTGAAGATCGCTTGGTGGCGGAGG 1531  
|||||  
482 yserGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyAsn 498  
1532 GGACGGTGCACCTGAATGCGGTAATCAATCCAGCCGACAACTCAT 1581  
|||||  
498 erThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIleIyr 514  
1582 TTCGGCTTTCGGCGGAGCGTGGATTGAAACGGGCAATTCGCTTCT 1631  
|||||  
515 PheGlyPheAlaGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531  
1632 CCACCGCATTCAAATACCGATGAAGGGCGATGATTCACACCAATC 1681  
| :|||  
531 easpHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnM 548  
1682 AAGACAAAGAATCCACCGTTACCATTTACAGCAATTAAGATTACT... 1728  
| :|||  
548 erThrAsnAlaSerAsnIleThrIleThrGlyGlnSerLeuIleThrAsp 564  
1728 ..... 1728  
565 ProAsnThrIleThrProTyAsnIleAspAlaProAspGluAsnPr 581  
1728 ..... 1728  
581 oTyAlaPheArgArgIleLysAspGlyGlyLeuLeuTyLeuAsnLeug 598  
1728 ..... 1728  
598 LuAsnTyThrTyThrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614  
1729 ..... ACAACGGCAATAC... 1743  
615 LeuProLysAsnSerGlyGlnSerAsnGluAsnThrLeuTyMetGly 631  
1744 ..... AACACTGGATGCAAAA 1762  
631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGlu 648  
1763 AAGAAATTCCTACACGGTTGGTTGGCGAAGAAATGCAACCAAAAGC 1812  
| :|||  
648 rGMetAsnGlyPheAsnGlyTyThrPheGlyGlnGlnGly...LysAsn 663  
1813 AACGGGGGCGCTCAATGCAATTTACACCGGAACGAGCGGATGCGACTTT 1862  
| :|||  
664 AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGlnGlnAsnArgH 680  
1863 ACTGCTTTCGGCGGAGCAAAATTTAAAGCAATATACGCAAAACAAAGC 1912  
| :|||  
680 eLeuLeuThrGlyIleThrAsnLeuAsnGlyAspLeuThrValGlnLysG 697  
1913 GCAAAATGTTTTCAGCGGAGACCGGACCGGACCGCTTCAATCATTTA 1962  
| :|||  
697 LyrThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713  
1963 GGAAGCGGGTGTCAAAATGAAGGT...ATCCACACAGAGAAAT 2006  
| :|||  
714 GlyLeuSerSerThrLysAspProHisPheAlaGlnAsnAsnGlnVal 730  
2007 CGTGTGGACACAGATTGATCGACCGCACATTTAAAGCGAAAGCTTC 2056  
| :|||  
730 lValValGluAspAspThrIleAsnArgAsnPheLysAlaThrThrMet 747

2057 ATATTCAAGGGGACAAAGCGTGGTTTC...CGCAATGTTGCCAAAGTG 2103  
| :|||  
747 snValThrGlyAsnAlaSerLeuTySerGlyArgAsnValAlaAsnIle 763  
2104 GAAGCGCATTTGCAATTAAGCAATCACGCCCAAGCAATTTTCGGTGGC 2153  
| :|||  
764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnHisIleGlyIyr... 779  
2154 ACCGCATCAAAAGCCACAAATCTGTACAGCTTGGAGATGAGCGGCTGA 2203  
| :|||  
780 ...LysThrGlyAspThrValCysValArgSerAspTyThrGlyIyr 795  
2204 CAATTTACCGGAAAAACCATTAACGAGATTAAGTGAATGTTGCTCAT 2253  
| :|||  
795 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810  
2254 AGCAAGACGACATCAGAGCAATGTGACGCTTGGCATCAGCTCATTT 2303  
| :|||  
811 AsnProThrAsnLeuArgLysAsnValAsnLeuThrGlnSerAlaAsnPh 827  
2304 AATTCACAGAGACTTGCACACTCAACGGCAATCTTAGTGCAGGCGAG 2353  
| :|||  
827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGly 843  
2354 ACACGCACTTACGGTTACCGCGCAACGCCCAACCAACGCAACTCAGC 2403  
| :|||  
843 snSerGlnValArgLeuThrGluAsn... 851  
2404 CTCGTGGCAATGCCCAAGCAACATTTAATCAGCCCAATTAAAGGCA 2453  
851 ..... 851  
2454 CACATGCGCTTGGACAAATGCTTCAATTATTAACCAACAGCGGTAC 2503  
851 ..... 851  
2504 AAAACGCACTGTGACGCTTTCGACACAGCTAAGCAACAGTAAGCCAT 2553  
851 ..... 851  
2554 TCCGACCTCAACGGCAATGTCCTCGTACCGATTAAGCAGATTTCATT 2603  
851 ..... 851  
2604 TGAAGACAGCGCTTACCGGAAATATCAGCGGCGCAAGCATACGGCAT 2653  
851 ..... 851  
2654 TACACTTAAAGACAGCAATGAGACGCTGCGTGGGACGGAATTAGGC 2703  
| :|||  
852 .....SerHisThrPheLeuThrGlyAsnSerAspValHis 863  
2704 AATTAAACCTTGACACGCCACCATTTACATGCAATTCGCTATCGACA 2753  
| :|||  
864 GlnLeuAspLeuAlaAsnGlnLysIleHisLeuAsnSerAlaAspAsn 880  
2754 CGATCGGAGGCGGCAAAACGGGCAATGCGGCAATGCGCGCGCGCC 2803  
880 rAsn... 881  
2804 GTTCGCGCGTTCCTATTATCGGTACCGCCCACTTCGCGAGAAATCC 2853  
| :|||  
882 .....AsnValThr... 884  
2854 CGTTTCAACAGCTGACGCTTAACCGCAATTTGAACGTCAGGGAATTT 2903  
| :|||  
885 LysTyAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPh 900  
2904 CGGCTTATGTCGAACTTTCGGCTACCGGACGCGCAATTTGAAGCTGG 2953  
| :|||  
900 eTyTyLeuThrAspLeuSerAsnLysGlnGlnLysPylValValIyr 917  
2954 CGGAAAGTTCCGAAGGCACTTACACTTGGCTGTGTCACAAATACGGCAAC 3003

```

917 hrrysSerAlaThrGlyAsnPhetHrLeuGlnValAlaAspLysThrGly 933
    :::::::::: :::::::::: ::::
3004 GAACCCGTAAGTCTGAGCAATATGACGGTAGTGGAAGAAAGAACACAC 3053
    ||||| ::::::::::::::: ::::
934 GluPro...AsnHisAsnGluThrLeuPheAspAla.....SerLys 947
    ::::::::::::::: ::::
3054 ACCGCTGTCCGAAATCTTAATTTACCTCGCAAAACGAAACACGGTAGT 3103
    ::::::::::::::: ::::
947 sAlaGlnArgAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
    ::::::::::::::: ::::
3104 CCGCGCATGCGCTATTCACCTTATCCGCAAGACGCGCATGTCGCGCTG 3153
    ||||| ::::::::::::::: ::::
964 euGlyAlaThrPlysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeu 980
    ::::::::::::::: ::::
3154 CATATATCCG..... 3162
    ::::::::::
981 TyrAsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIl 997
    ::::::::::::::: ::::
3162 ..... 3162
997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
    ::::::::::::::: ::::
3162 ..... 3162
1014 snGluGluIleAlaArgValAspGluAlaProValProProAlaPro 1030
    ::::::::::::::: ::::
3162 ..... 3162
1031 AlaThrProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGl 1047
    ||||| ::::::::::::::: ::::
3162 ..... 3162
1047 userLysThrValGluLysAsnGluAsnAlaThrGluThrThrAlaG 1064
    ::::::::::::::: ::::
3163 ..... 3177
    GTCAAAGAACACAGAG
1064 InAsnArgGluValAlaLysGluAlaLysSerAsnValLysAlaAsnThr 1080
    ||||| ::::
3178 CTTCGCAACAACTCGGCAGGCGGGA...GAACAGAGCGCGCTTGAC 3224
    ||||| ::::::::::::::: ::::
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097
    ::::::::::::::: ::::
3225 GGCAGAAACAGGACACACTGCGCGCAACACAGCGCGAAAGACACAG 3274
    :::::::::: ::::
1097 rThrGluThrLysGluThrAlaThrValGluLysGluLysAlaLysV 1114
    :::::::::: ::::
3275 CGCAAGC.....CTTGACGCGCTGATTCG 3300
    :::::::::: ::::
1114 alGluThrGluLysThrGlnGluValProLysValThrSerGlnValSer 1130
    :::::::::: ::::
3301 GCGCGGCGCAATGCCACGAAAGAGCAAAAGTTGCGCGAGCGCGCG 3350
    :::::::::: ::::
1131 ProLysGlnGluGlnSerGluThrValGlnProGlnAlaGluProAlaAr 1147
    :::::::::: ::::
3351 GCAGGAGGCGCGGAAATGCCGCAATATGACGCGAGAGAGAGAA 3400
    ||||| ::::
1147 q.....GluAsnAspProThrValAsnIleLysGlnProLys 1160
    :::::::::: ::::
3401 AACGGGTGACGCGGATTAAGACACCGCTTG...GGAAACAGCGGAA 3447
    :::::::::: ::::
1160 eArgGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176
    :::::::::: ::::
3448 GCG..... 3450
    ::::
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
    :::::::::: ::::
3451 .....GAACCCGCGCGCTACACACCGCTTCC 3478
    ::::||| |||||
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
    ::::||| |||||
3479 CC.....CGCGCGCGCGCG 3492
    || |||||
```

```

1209 roThrValAsnSerLysSerAsnLysProLysAsnArgHisArgArg 1225
    ::::
3493 GCCCGCGCGGATTTGCCG...CAACCCAGCCCCCAACCCGCAACCC 3539
    :::: |||||
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
    ::::
3540 GCAGGCG.....GACCTGATACGCGTTATGCCATA 3571
    :::: |||||
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259
    :::: |||||
3572 GCGGTTTACGATTTTCCGCG.....ACGCTAACAGC 3606
    :::: |||||
1259 lAvalLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
    :::: |||||
3607 GTTTCGCGCGTACAGACGAAATGAGCCGCGT...TTTGCCGAAGACCG 3653
    ||||| ::::
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGlnLeuAsnAsnGluGl 1292
    ::::
3654 CCGCAAGCCGCTTTGGACACGCGCATCCGGGACACCAACACTACCGTT 3703
    :::: |||||
1292 yGlnTyrAsnValTyrValSerAsnThrSerMetAsnLysAsnTyrSerS 1309
    :::: |||||
3704 CGCAGATTTCCGCGCTACCGCCCAACCAACCGACTCGGCCAATCGGT 3753
    ||||| ::::
1309 eSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
    :::: |||||
3754 ATGCAGAAAACCTCGGACGCGG...CGGCTCGGCATCTGTTTGCA 3800
    :::: |||||
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
    :::: |||||
3801 CAACCGGACCGGAACACCTTCGACAGAGCGCATCGGCAACCTCGGACGCG 3850
    :::: |||||
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrLeuA 1359
    :::: |||||
3851 TTGCCACGCGCTTTCGGGCAATACGCGCATCGCGAGGTTGACATC 3900
    :::: |||||
1359 lAglnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTyrPylLeu 1375
    :::: |||||
3901 GGCATCAGCGCGCGCGGCTTTAGTACGCGGACCTTCACAGCGCAT 3950
    ||||| ::::
1376 GlyIleAspLeuGlyTyrGlyLysPheGlnSerLysGlnThrAsnHis 1392
    :::: |||||
3951 CAGAGGCAAAATCCCGCGCGCTGCTGATTCAGGCAATTCAGGCAAT 4000
    :::: |||||
1392 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409
    :::: |||||
4001 ACCGCGCAGGTTTCGGCGGATTCGCGATCGAAGCCGACATCGCGCGCA 4050
    :::: |||||
1409 yAlaIleAsnHisLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1425
    :::: |||||
4051 CGCTATTTGCTCCAAAACCGGATTAACGATACGAAACGTCATATATGCG 4100
    :::: |||||
1426 TyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIleL 1442
    :::: |||||
4101 CACCCCGGCGCTTGCAATTCACCGGCTACCGCGCGGATTAAGCAGATT 4150
    :::: |||||
1442 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerT 1459
    :::: |||||
4151 ATTCAATTCACACCGCGCAACACATTTCCATCAGCGCTTATTTGAGCGTG 4200
    :::: |||||
1459 yThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1474
    :::: |||||
4201 TCCTATACGATGCGCTTCGCGAAAGTCCGACGCGCGTCAATACCGC 4250
    ||||| ::::
1475 ArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGlyTy 1490
    :::: |||||
4251 CGTATTCGCGCAGATTTGGCAAAACCGCGAGTGCGGAATGGCGGATA 4300
    :::: |||||
1490 rAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuL 1507
    :::: |||||
4301 ACCCGCAAAATCAAGGTTTCAGCGTGTCCCTCCACGCTCCCGCGCAAG 4350
    :::: |||||
1507 yLeuLysTyrHisAsnValLysLeuSerLeuIleGlyLeuThrLys 1523
    :::: |||||
```



4351 GGCCCGCATTTGGAGCGCAGACGCGCGCATCAATTAAGGCTAC 4398  
 1524 AAlayGlnAlaGlnLysGlnLysThrAlaGlnLysLeuSerPhe 1539

seq\_name: /SID1/9cdata/geneseq/geneseq-emb1/AA2000.DAT:AA75564

seq\_documentation\_block:  
 ID AA75564 standard; Protein: 1431 AA.

AC AA75564;  
 DT 21-MAR-2000 (first entry)  
 DE Neisseria meningitidis ORF 759 protein sequence SEQ ID NO:2602.  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy.  
 OS Neisseria meningitidis.  
 PN WO957280-A2.  
 PD 11-NOV-1999.  
 PE 30-APR-1999; 99WO-US09346.  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094669.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 PA (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 PI Fraser C, Galatelli C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 DR WPI: 2000-062150/05.  
 DR N-PSDB; AA254326.  
 PS Claim 2: Page 1234; 1453pp; English.  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 SQ Sequence 1431 AA;

alignment\_scores:  
 Quality: 1585.50 Length: 1668  
 Ratio: 1.829 Gaps: 50  
 Percent Similarity: 51.978 Percent Identity: 28.118

alignment\_block:  
 US-09-303-518D-653 x AA75564 ..

Align seg 1/1 to: AA75564 from: 1 to: 1431

55 GGCCGCATCCGCTTCTGCGCCGCTTACTTACCATATGCTGCTGG 104  
 17 GlnLeuPheAlaValSerProAlaTyrSerSerIleValArgasn ..... 31  
 105 CATCTGCGCCCAAGCCCGGGGAGACACACTTATTGGCATCACTACC 154  
 32 ..... AspyAlaAspyArg 36  
 155 AATACATTCGGGACTTTCGCGAATAAAGCAAGTTGCGATCGGGCG 204  
 36 IntyrPheArgAspPheAlaGlnLysnLysGlyAlaPheThrValGlyAla 52  
 205 AAAGATATGAGCTTACAAACAAAAAGGGAGTGGTCCGCAATCGAT 254  
 53 SerasnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValle 69  
 255 GACGAAGCCCGATGATGATTTCTGTGATCGCT...AACGGCG 301  
 69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86  
 302 TGGCGGCAATGGCGCGCGCATCATATATTTGTAGCGTGACATTAACGC 351  
 86 IeAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102  
 352 GGCTATACATGATGTTGTTGGTGGGAGGAGCAAGCATCCGATCAGCA 401  
 103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProIuGluG 119  
 402 CCGCTTTCTTACCAATGTGAAAGAAATAATTAATAAGCAGGAGCTA 451  
 119 nAlaTyrThrTyrArgLeuValSerArg ..... A 129  
 452 ACGGCATCCTTATGGCGGATATCATATGCGCGGTTTGGCAATTT 501  
 129 snProHisPro...AspTyrAspTyrHisLeuProAlaLeuAsnLysLeu 144  
 502 GTCACAGATCAGACACCTTGGATGACACAT ..... 534  
 145 ValThrGlnLysSerProThrAlaLeuSerSerValProLeuLeuLys 161  
 535 ....TATATGATGGGTGGAAATACGCTGATTAATAATACCTGATC 580  
 161 nGlyGlnProLysAlaAsnAlaTyrLeuAspThrAspArgPheProTyrP 178  
 581 GTGTTGAATCGGAGCGAGCAGACATATTTGGCGTGTGATGAGCGAA 630  
 178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193  
 631 CCCAATTAACCGCAAGATTCATATCATATTTGCAAGGCAATATCTTGCT 680  
 194 .....ThrArgThrArgThrAlaProAlaTyrGlyLeu 205  
 681 CGTGGTGCAATACCTTTGACAAATAAGATCAGGTGGTGCACAGTGA 730  
 205 uThrGlyGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeu 222  
 731 ACTTAGTACCGCAAAATAATTAACATAGCCATATGTTTATACCAACA 780  
 222 euValGlyGly...SerLeuThrAspGlnPro.....LeuAsnThr 234  
 781 GAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCA 830  
 235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysH 251  
 831 AAAGCAAAAGTGTATATGATGGGTATTTGCAACAGCAACCCCTATA 880  
 251 sGluAsnArgIleThrValLeuAlaGlyValLeuSerThr.....TyrA 265



813 ..... ||| ::|||  
GlySalahis..... 816  
2368 GTTACGGCAAGCCCAACCAACGCAACCTCACCTCGTGGCAATGEC 2417  
816 .....  
2418 CCAAGCAACATTTAATCAAGCCACATTAAGCGCAACACATCGCTTGGG 2467  
816 .....  
2468 ACAATGCTTATTATTAAGCAACACCGCGTACAAACGCGAGCTGG 2517  
816 .....  
2518 AGCGTTTCGACACAGCTAAGGCAACGTAACCATTCGCGACTCAAGG 2567  
816 .....  
2568 CAATGCTCCCTAGCGCATTAAGCAGATTCATTTTGAACACCGCGT 2617  
817 ..... L 817  
2618 TTACCGCAAAATTCAGCGCGCAAGATACGGCATTAACATAAA... 2664  
817 euryglyserileargylasprthalaivalargmetgluala 833  
2665 GACACGCAATGAGCGCTGCGGCGCAAGCAATTAAGCAATTTAAACT 2714  
834 AspserasntrpThrLeuSerGlnSerHisThrGlyAlaLeuThrLe 850  
2715 TGACAAGCGCCACCATTAACATCAATTCGCGTATGACACAGATGGCGAG 2764  
850 uasprglyalaglnilethrLeuasnProAspPheAlaIasnasn... 864  
2765 GCGGCGCAACCGCGAGTGGCGAGATGGCGCGCGCGCGTTCGCGCGCT 2814  
864 .....  
2815 TCCTATTATCCGTTACCGCGCAACTTCGCGAGATCCGTTTCANAC 2864  
865 ..... ThrHisasnAsnArgPheAsnTh 872  
2865 GCTGACGTAACGCAAAATGAGCGGTACAGGAACATTCGCTTATGT 2914  
872 rleuthrvalasnglythrLeuaspGlyPheGlyThrPheArgPheLeuT 889  
2915 CGGAACCTTCGCGTACCGCAAGCGG... AATGGAACCTTCGCGGAAGT 2961  
889 hrGlyIleValArgGlyGlnAsnAlaProPheLeuLysLeuGlyasp 905  
2962 TCAGAGGCACTTACACCTTGCTGTCAACATACCGCAAGCAAGCCGT 3011  
906 SerArgGlyAlaPheGlnIleHisValLysasnThrGlyGlnIuProG 922  
3012 AAGTTCAGCAAAATGAGCGGTAGGGAAGAAACACACACCGCGCT 3061  
922 nThrThrGlnSerLeuAlaLeuVal... SerLeuasnProLysHis 937  
3062 CGGAATACTTAATTCACCTGCAAAACGAACAGTCGATGCGGCGCA 3111  
937 erHisGlnAlaArgPheThrLeuGlnAsnGlyTyrAlaAspLeuGlyAla 953  
3112 TGGCGTTATCAGCTTATCCGCAAGAGCGCGAGTTCGCGTCAATATCC 3161  
954 TyrArgTyrIleLeuArgLysAsnAsnGlyTyrSerLeuTyrAsnPr 970  
3162 GGTCAAGAAACAGAGCTT... 3180  
970 OleuLysGlnAlaGlnLeuGlnIleGlnAlaThrArgAlaGlnHisGln 987  
3180 ..... 3180  
987 rGAsnGlnGlnAlaTyrAsnGlnLeuGlnAlaThrAspIleSerArgGln 1003  
3181 ..... TCAGCAAACTGGCGAAGCG... 3201  
1004 ValGlnHisAspSerAspAlaThrArgGlnAlaLeuGlnAlaTrpGlnAs 1020  
3202 .GGCAAAACAGAGCGCGCTTGACGCGCAAAACAGCAACTGCGCCCA 3250  
1020 nSerGlnThrGlnLeuAlaArgIleAspSerGlnValGlnTyrLeuSerA 1037  
3251 AACACAGCGCGAAAAAGAC... AACGCGCA 3279  
1037 laglnleuLysGlnThrAspProLeuThrGlyIleLeuThrArgAlaGln 1053  
3280 AGCCTTGACGCGCTGATTCGCGCGCGCGCGCAATGCC... 3315  
1054 AsnLeuLysAla... AlaGlnGlyTyrSerAlaAspIleLecysArgI 1068  
3315 ..... 3315  
1068 nValAlaLysAlaAlaAspThrAsnAspLeuThrLeuPheGlnThrGln 1085  
3316 ..... ACCGAAAAAGCGAAGAGTTCGCGCAACCG 3345  
1085 euAspThrTyrIleGlnArgValGlnMetAlaGlnSerGlnLeuAspLys 1101  
3346 GCGCGCGACGACGCGCGGGAATGCCGCAATATGACGCGGAGAGACA 3395  
1102 AlaArgGlnGlyGlyasp... AlaGlnAlaValGlnThr 1113  
3396 GAAAAAAGGCGTGCAGCGGATTAAGACACCGCGCTTCGCGCAACAG...C 3442  
1113 rAlaArgHisAlaTyrLeuAsnAlaLeuAsnAlaGlyLeuSerArgGlnIleH 1130  
3443 GCGAAGCGGAACCGCGCGCTACACACCGCTTCGCGCGCGCGCGCG 3492  
1130 lsserLeuLysThrGlyValAlaGlyTylargmetProAsnLeu... 1144  
3493 GCCCGCGCGGATTTGCCGCAACCGCAAGCCCAACCGCAACCCACCGCA 3542  
1144 ..... 1144  
3543 GCGGACCTGATACGCCGTTATGCAATAGCGGTTGAGTAATTTCCG 3592  
1145 .AlaGlnLeuIleSerArgSerAlaAsnThrAlaValSerGlnAlaAla 1161  
3593 CCACGCTCAACAGCGTTTCGCGGTACAGGCAATTCGACCGCGGTGT 3642  
1161 lAtyrAsnThrGlyArgGlnGlnAlaGlyArgArgIleAspArgHisLeu 1177  
3643 GCGAAGACCGCGCAACCGCGCTTGCAACAGGCGCATCGGGAACCA 3692  
1178 ThrAspProGlnGlnAsnIleThrLeuGlnThrGlyThrGlnIleThr 1194  
3693 ACACTACCGTTCCGAAGATTCGCGCGCTACCGCAACCAACAGCATGCG 3742  
1194 rAspTyrHisSerGlyThrHisArgProTyrGlnGlnIleThrHisTyrAla 1211  
3743 GCCAATCGGTATGCAAAAAACCTCGGCAAGCGG... CGCGTGGCAATC 3789  
1211 lAhisIleGlyIleGlnThrGlyIleThrAspArgLeuSerValGlyThr 1227  
3790 CTGTTTTCGCAACCGCGCAACCGCAACCTTCGAGACGCGGCGG... 3837  
1228 lIleLeuThrAspLeuArgThrHisAsnAlaArgPheAspGlnGlyAlaSerAl 1244  
3838 .... AACTGCGACGGCTTCGCGACGCGTTCGCGCAATACGCA 3883  
1244 aArgAsnArgSerAsnGlyAlaHisLeuPheValLysGlyGlnAsnGlyAla 1261  
3884 TCGGCAAGTTTCGACATCGGCATCGAGCGCGCGCGGTTTGTAGCGCG 3933  
1261 la.....LeuPheAlaAlaAlaAspLeuGlyTyrSerAsnSer 1273



```
442 GCAGGACTAAAGCCATCTTATGCGGATATATCATATGCGCGTT 491
   ::::::::::::::::::::::::::::::::::::::::::::
156 Ser.....SeraspLeuHisThrProVal 164
492 GCACAATTTGTCACAGATGCAAGCCTGTGAGATGACAGTATATAG 541
   |||||::::::::::::::::::::::::::
164 uAspLysLeuValThrGluValAlaProAlaThrValThrSer..... 179
542 ATGGGTGAATACGTCGATTTA.....AATAATACCGTCGATCCT 582
   ::::::::::::::::::::::::::::::::::::::::::::
180 .....SerThrAlaSpLLeuAsnProSerLysTyrSerAlaPhe 193
583 GTTGCAATCGGACGACAGCAATATTTGCGGTCTGATGAAAGCAAGC 632
   |||||::::::::::::::::::::::::::
194 TyrTrpGlnGlySerGlnTyrLLeuLAsnSerGlnGlyLys.. 209
633 CAATACCGCGAAGTTCATATCATATTTGCAAGCGCATATTTCTGGCTG 682
   ::::::::::::::::::::::::::::::::::::::::::::
210 .....AlaHisThrValThrGlyLysTyrGlyLysLeu 221
683 TCGGTGGC.....ATACCTTGCACAAATGATCAGTGTGCGC 723
   ::::::::::::::::::::::::::::::::::::::::::::
221 hTrpGlyLysLeuProThrSerPhePheTyrHisGlySerAspGlyLe 237
724 ACAGTCAACTTAGCTAGCAGAAATAATTAACATAGCCCATATGTTT 773
   ::::::::::::::::::::::::::::::::::::::::::::
238 GlnLeuTyrMetGlyLysAlaHisLeuHisAspHisSer.....Ile 251
774 ACACAACAGGAGCTCATTTGGCGACAGTGGCTGCAATGTTATTCATG 823
   |||||::::::::::::::::::::::::::
251 uProSerPheGlyLysAlaGlyAspSerGlySerProLeuPheGlyTrp 268
824 ATGCCCAAGCAAGTGGTATATTAATGGGTATTCACAACAGCAAC 873
   ::::::::::::::::::::::::::::::::::::::::::::
268 snrAlaLysGlyLysTrpGlnLeuValGlyVal..... 279
874 CCTATATAGAAAAAGCAATGGCTTCAGCTAGTGTGTA 915
   |||||::::::::::::::::::::::::::
280 ...TyrSerGlyValGlyGlyLysThrAsnLeuLysSerLeuIlePr 295
916 GATGGTCTCATGATCAATCTTGGCTGAGATACCCATTCAGTATCT 964
   ::::::::::::::::::::::::::::::::::::::::::::
295 ogLAsnSerPheLeuSerGlnLysSerGluAsp..... 306
965 AGCAACCATCAAAATGGAAATACCTTTTAAAC..... 999
307 .....AsnAspAlaProValPhePheAsnAlaSerSerGlyAla 319
1000 .....GACAATATATATGCGGACAGAAATAATCGATGC 1031
320 ProLeuGlnTrpLysPheAspSerSerThrGlyThrCly..... 332
1032 CAACAATAAACATATCTTACCTTATAGATTAATAACAGAACCGTTC 1081
332 ..... 332
1082 AATTGTTTATGTTCTTATTCGAGACAGCAGACACTGTTATCAT 1131
   |||||::::::::::::::::::::::::::
333 .....SerLeuLysGlnGlySerAspGlnTyrAlaMetHis 344
1132 GCTGACAGTGGGTCAACAGTATTCGACCGACAGTCAATATGATGAAA 1181
   ::::::::::::::::::::::::::::::::::::::::::::
345 GlyLysGlySerAsp.....LeuAsnAlaGlyLysAs 356
1182 TATTTCTTATGACAAAGAAAGAGTGAATGATCTTACAGCAACA 1231
   ::::::::::::::::::::::::::::::::::::::::::::
356 nLeuThrPheLeuGlyHis...AsnGlyLysLLeuAspLeuLysAsnSer 372
1232 TCACAACAGCGCGGCGGTTGTTATTTGAGGGTAAATTTTACGCTCG 1281
   ::::::::::::::::::::::::::::::::::::::::::::
372 aLThrGlnGlyAlaGlySerLeuThrPheThrAspAspTyrThrValThr 388

1282 CTTAAAAACAGCAAACTGTGCAAGCGCGGCGTTTCATATCATGATG 1331
   ::::::::::::::::::::::::::::::::::::::::::::
389 ThrSerAsnLysSerThrTrpThrGlyAlaGlyLysLeuValAspLysAs 405
1332 CAGTACCGTTACTTGGAAAGTAAGCGCGTGGCAAGCAAGCGCTGTCCA 1381
   ::::::::::::::::::::::::::::::::::::::::::::
405 pAlaSerValAsnTrpGlnValAsnGlyValLysGlyAspAsnLeuHisL 422
1382 AAATCGGCAAAAGCGACGCTGTGTTCAAGCCAAAGGGGAAAAACAAGC 1431
   ::::::::::::::::::::::::::::::::::::::::::::
422 yLLeuGlyGlnGlyLysLeuValValGlnGlyThrGlyValAsnGlyLys 438
1432 TCGGTGAGGCGGCGGAGTAAAGTCATCTTATGATCAGACGCGGACA 1481
   ::::::::::::::::::::::::::::::::::::::::::::
439 GlyLeuLysValGlyAspGlyThrValValLeuAsnGlnGlnAlaAspSe 455
1482 TCAGGCAAAAAACAGCCTTTAGTGAATCGCGCTGTGTCAGCGGAGG 1531
   ::::::::::::::::::::::::::::::::::::::::::::
455 rSerGlyHisValGlnAlaPheSerSerValAsnLLeuAlaSerGlyAsp 472
1532 GACGCTGCACTGAATGCCGATATCATGTCACCCGCAAACTCTAT 1581
   ::::::::::::::::::::::::::::::::::::::::::::
472 roThrValValLeuAlaAspAsnGlnGlnValAlaAsnProAspAsnLLe 488
1582 TTCGCGCTTCGCGGCGAGCTTGGATTGACGCGCATTCCTTTCGT 1631
   ::::::::::::::::::::::::::::::::::::::::::::
489 TrpGlyTyrArgGlyLysValLeuAspValAsnGlyAsnAspLeuThrPh 505
1632 CCACCGCATTCAAATACCGATCAAGGCGGATGATTCACACCAATTC 1681
   |||||::::::::::::::::::::::::::
505 eHisLysLeuAsnAlaAlaAspGlyLysAlaThrLeuGlyAsn...Ser 521
1682 AAGACAAGAATTCACCGTTACCAT.....ACAGCAATAAGAT 1722
   ::::::::::::::::::::::::::::::::::::::::::::
521 exAspLysThrAlaAsnLLeuThrLeuAspTyrGlnThrArgProAlaAsp 537
1723 ATTACT.....ACAACGCGCATTAACAACAACTTGATGATGCA 1760
   ::::::::::::::::::::::::::::::::::::::::::::
538 ValLysValAsnGlnLysTrpSerSerAsnArgGlyThrValGlySerLe 554
1761 AAAAGAAATGGCTACAC..... 1779
554 u.....TyrLLeuTyrAsnAsnProTyrThrHisThrValAspTyrPhe 569
1780 .....GGTGTGTT.....GGCGAAGAT... 1800
569 LeuLysThrSerSerTyrGlyTyrPheProThrGlyGlnValSerAsn 585
1800 ..... 1800
586 GlnHisTrpGlyTyrValGlyHisAspGlnAsnSerAlaGlnAlaLeuLe 602
1801 .GCAACCAAAACGAAC.....GGGCGGCTC... 1824
   ::::::::::::::::::::::::::::::::::::::::::::
602 uAlaAsnArgLLeuAsnAsnLysGlyTyrLeuLysHisGlyLysLeuLeu 619
1825 .AATCTGAATTCACACCGGAGACCGGATCGCACT.....TTA 1863
   |||||::::::::::::::::::::::::::
619 LysAsnLLeuAsnPheSerAsnLysAlaThrProGlyTyrThrGlyValLeu 635
1864 CTGCTTCGCGGCGGAAACAATTAAGCAATATCAGCAAAACAACAGG 1913
   ::::::::::::::::::::::::::::::::::::::::::::
636 ValMetAspGlySerAlaAsnMetSerGlyThrThrPheGlnGlnLysGln 652
1914 CAACCTGTTTTACGCGGACAGCAGCAGCCAGCC..... 1950
   ::::::::::::::::::::::::::::::::::::::::::::
652 yArgLeuThrThrLysGlnGlyHisProValLLeuHisAlaSerThrSerGln 669
1951 .....TACAATCATTTAGAGACGCGGTGTCAAAAATGGA 1986
   ::::::::::::::::::::::::::::::::::::::::::::
669 eTrlAlaAsnThrValSerSerLeuGlyAspAsn.....Ser 681
1987 GGTATCCCAAGAGAAATCGTGTGGACAACAAGATTGATGACCGGAC 2036
```

```
682 ValLeuThrIleProThrSerPheThrGlnAspSPTripGluAsnArgTh 698
      :::: ||| :::: ||||| :::: |||||
2037 ATTTAAAGCGGAAACTTCATATTCAGGGCGGCAAGCGGTGTTCC 2086
      ||||| :::: :::: ::::
698 rPheSerPheGlySerLeuValLeuLysAspThrAspPheGlyLeuGlyA 715
      :::: :::: ::::
2087 GCAATGTT.....GCCAAAGTGGAAAGCGCATTCGCAATTAAGCAAT 2127
      ||||| :::: ::::
715 rGlnAlaThrLeuAsnThrThrIleGlnAlaAsp..... 726
      ||||| :::: ::::
2128 CACGCCCAAGACAGTTTCGGTGTGCGACCCGATCAAGCCACACAATCTG 2177
      :::: :::: ::::
727 AsnSerSerValThrLeuGly.....AspSerArgValPheIle 739
      :::: :::: ::::
2178 TACACGTTCCGACTGACGGGTCTGACAAAGTTGACGAAAAACCATTA 2227
      :::: :::: ::::
739 eAspLysLysAspLysGlnGly.....ThrAlaPheThrLeuG 752
      :::: :::: ::::
2228 CCGACGATAAGTATGCTTCA.....TTGACGACGCGACATCGA 2271
      :::: :::: ::::
752 LuGluGlyThrSerValAlaThrLysAspAlaAspLysSerValPheAsn 768
      :::: :::: ::::
2272 GCGAATGTCAAGCTTCGCCGATGACGCTCATTTAAATCTCAGAGACTTGC 2321
      ||||| :::: ::::
769 GlyThrValAsnLeuAspAsnGlnSerValLeuAsnIleAsnGluIle.. 784
      :::: :::: ::::
2322 CACACTACAGCGCAATCTTAGTGCAGCGGAGACAG..... 2358
      :::: :::: ::::
785 ....PheAsnGlyIleGlnAlaAsnAsnSerThrValAsnIleSers 800
      :::: :::: ::::
2359 .....CACTATACGGTTACGCGC.....AAC 2379
      :::: :::: ::::
800 eAspSerAlaValLeuGluAsnSerThrLeuThrSerThrAlaLeuAsn 816
      :::: :::: ::::
2380 GCGACCCAAAGCGCAACTC.....AGCTCTGGGCAATGC 2417
      :::: :::: ::::
817 LeuAsnLysGlyAlaAsnValLeuAlaSerGlnSerPheValSerAspG 833
      :::: :::: ::::
2418 CCAAGCAACATTTAAATCAAGCGCATTAAGCGCAACATCGGCTCGG 2467
      :::: :::: ::::
833 yProValAsnIleSerAspAlaThrLeuSerLeuAsn...SerArgPro 849
      :::: :::: ::::
2468 ACAATGCTTCATTTAATCTAAGCAACAAGCGGCTTACAAAAGCGCACTCG 2517
      ||||| :::: ::::
849 sPGLuValSerHisThrLeuLeuProValLysArgLysGlySerThr 865
      :::: :::: ::::
2518 AGCGTT...TCCGACACGCTAAGCGCAAGCTAAGC...CATTCGCGACT 2561
      :::: :::: ::::
866 AsnLeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLe 882
      :::: :::: ::::
2562 CAGCGGCAATGCTCTCCCTAGCGCGATAG..... 2589
      :::: :::: ::::
882 uSerGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyLysG 899
      :::: :::: ::::
2590 .....GCA 2592
      :::: :::: ::::
899 LuGlyGluLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSer 915
      :::: :::: ::::
2593 GTATTCCTTTTGAACACGCCCTTACCGGAAAAATCAGCGCGGCA 2642
      :::: :::: ::::
916 LeuPheAsnGlyTyrArgAsnThrTrpSerGlySerLeuAsn...AlaPr 931
      :::: :::: ::::
2643 GGATACGCGCATTTACATTAAGACAGAGAAATGAGCGTCCGCGGCA 2692
      ||||| :::: ::::
931 oAspAlaThrValSerMetThrAspThrGlnTrpSerMetAsnGlnSns 948
      :::: :::: ::::
2693 CGGAATTAGCGCAATTAAACCTTGACAAAGCGCACCATTTACACTCAATTC 2742
      :::: :::: ::::
948 eThrAlaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsn... 963
      :::: :::: ::::
2743 GCCTATCGACACGATGCGGCGGCGGCAACCGGCGACTGCGGCGATGC 2792
      :::: :::: ::::
963 ..... 963
2793 GCGCGCGCGCGCTGCGCGCGCTCCCTATTATCCGTTACGCCGCAACT 2842
      :::: :::: ::::
964 .....G 964
2843 CGGCAAAATCCCGTTTCAACAGCTGACGCTAAACGGCAAAATGACGCT 2892
      :::: :::: ::::
964 LyGlyThrSerSerPheThrThrLeuThrThrAspAsnLeuAspAlaVal 980
      :::: :::: ::::
2893 CAGGACATTCCTCGCTTAATGTCGGAAGCTTCGGCTACCCGACGCGCA 2942
      ||||| :::: ::::
981 GlnSerAlaPheValMetLargThrAspLeu.....AsnLysAlaAspLys 995
      :::: :::: ::::
2943 ATTGAAGCTGGCGGAAAGTCCGACAGC...ACTTACACTTGGCTGCA 2989
      ||||| :::: ::::
995 sLeuValIleAsnLysSerAlaThrGlyHisAspAsnSerIleTrpValA 1012
      :::: :::: ::::
2990 AC.....AATACCGGCAAGCAACCC...GTA 3012
      :::: :::: ::::
1012 snPheLeuLysLysProSerAspLysAspThrLeuAspIleProLeuVal 1028
      :::: :::: ::::
3013 ACTCTCGACCAATTGAC.....GT 3032
      :::: :::: ::::
1029 SerAlaProGluAlaThrAlaAspAsnLeuPheArgAlaSerThrArgVa 1045
      :::: :::: ::::
3033 AGTGAAGGAAAGACAAACACACCGCTGTCGAAATCTTATTTCACCC 3082
      ||||| :::: ::::
1045 lValGlyPheSerAspValThrProThrLeuSerValArgLysGluAspG 1062
      :::: :::: ::::
3083 TCGAAAACGACACAGCTGCATGCCGCGCATGCGGTATACGCTTATCCG 3132
      :::: :::: ::::
1062 lLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1075
      :::: :::: ::::
3133 AAAGACGGCGAG.....TTCGCGCTGCA 3155
      :::: :::: ::::
1076 AsnAspGlyGlnGlyLysAlaAlaAlaThrPheMetHisIleSerTyrAs 1092
      :::: :::: ::::
3156 TAATCCGCTCAAAAGA.....CAAGAGCTTT 3181
      :::: :::: ::::
1092 nasPheIleThrGlnValAlaAsnAsnLeuAsnLysArgMetGlyAspLeuA 1109
      :::: :::: ::::
3182 CCGACAACTCGGCAAGCGCGGAGAAACAGAGCCGCTTACGCGCAAA 3231
      :::: :::: ::::
1109 rGAspIleAsnGlyLysLysLysLysLysLysLysLysLysLysLysLys 1125
      :::: :::: ::::
3232 CAGGCA.....CAACTGC 3245
      :::: :::: ::::
1126 GlySerAlaAspLysLysLysPheThrAspHisTyrThrLeuLeuGlnMetG 1142
      :::: :::: ::::
3246 CCGCAAAACAGCGCGGAAAGACACAGCGCAAAAGCTTGAACGCGCTGA 3295
      :::: :::: ::::
1142 yAlaAspArgLysHisGlyLeuGlySerMetAspLeuPheThrGlyValM 1159
      :::: :::: ::::
3296 TTCGCGCGCGCGCAATGCCACCGCAAAAGCGCAAAAGTTTCCCGACCG 3345
      :::: :::: ::::
1159 eTAlaThrTyrThrAspThrAspAlaSerAlaGlyLeuTyrSerGlyLys 1175
      :::: :::: ::::
3346 GCGCGGCGGAGCGCGGAAATGCGGCAATTAATGCAAGCGGAGAGAGA 3395
      :::: :::: ::::
1176 ThrLysSerThrLysLysLysLysLysLysLysLysLysLysLysLysLys 1189
      :::: :::: ::::
3396 GAAAAAAGCGGTGCAGCGGATTAAGACACCGCGCTTGGGCAAAACAGCGG 3445
      :::: :::: ::::
1189 eArgSerGlyAlaTyrPheAsp.....LeuIleAlaLysTyrIleH 1203
      :::: :::: ::::
3446 AACGGAACCGCGCGCGGTACACCGCTTCCCGCGCGCGCGCGCGGCG 3495
      :::: :::: ::::
1203 lAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1207
      :::: :::: ::::
3496 CGCGGGGATTTGCCGCAACGCGAGCCCAACCGCAACCGCAACCGAGCGG 3545
      :::: :::: ::::
1208 .....Ty 1208
```







```

694 hrLeuSerLeuysaspAlaIaspheHisLeuSerArgAsnAlaSerLeu 710
      ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
2140 GTTTCGGGTGCGCACCGCATCAAGCCAC..... 2169
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
711 AsnSerAspIleYsSerAspAsnSerHisIleThrLeuGlySerAspAr 727
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2170 .ACATCTGTACACGTTCCGACCTGACGGGT...CTGACAGTGTGTACCG 2215
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
727 gValPheValaspIysAsnAspGlyThrGlyAsnTyValIleLeuGluG 744
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2216 AAAAACCATTAACCGACGATAAAGTATTCCTCATTCAGCAAGACCGAC 2265
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
744 IuGlyThrSerValProAspThrVal.....AsnAspArgSerGln 757
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2266 ATCAGAGCATATGTCACGCTTCGCGATCAGCGCTATTAATCTC... 2310
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
758 TyrGluGlyAsnIleThrLeuAspHisAsnSerThrLeuAspIleGlySe 774
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2311 .....ACAGCA..... 2316
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
774 ArgPheThrGlyGlyIleGluAlaTyrAspSerAlaValSerIleThrS 791
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2316 ..... 2316
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
791 erProAspValLeuLeuThrAlaProGlyAlaPheAlaGlySerSerLeu 807
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2317 .....CTGCGCACTCAACCGCATCTTAATCTTACTGC 2345
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
808 ThrValHisAspGlyGlyHisIleThrAlaLeuAsnGlyLeuPheSer.. 823
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2346 AGCGGAGACACGCACTATACGCTTACCGCAACGCCCAAAAGCGCA 2395
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
824 .....AspGlyHisIle.....GlnAlaGlyLysAsnSerI 834
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2396 ACCTCAGCCTCGTGCGCAAT...GCCACAGCATTTAATCAAGCCACA 2442
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
834 ySleIleThrLeuSerGlyThrProValIysAspThrAlaAsnGlnTyAla 850
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2443 .....TTAAACGGCAACACATCGGCTTCGACATGCTTC 2477
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
851 ProAlaValTyrIleuThrAspGlyTyrAspLeuThrGlyAspAsnAlaTh 867
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2478 ATTAACTAAGCAACAACGCC..... 2499
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
867 rLeuGluIleThrArgGlyAlaHisAlaSerGlyAspIleHisAlaSerA 884
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2500 .....GTCAAAACGGCGAGTGTGACGCTTCCGACACGCTAAG 2538
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
884 lAlaIaSerThrValThrIleGlySerAspThrProAlaGluLeuAlaSer 900
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2539 GCAAAACGTAAAGCATTCGCACTCAACGGCAATGTCCTAGCCGATAA 2588
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
901 AlaGluThrThrAlaSerAlaPheAlaGly..... 910
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2589 GCGAGTATTCATTTTAAACAGACCGCTTACCGGAAAAATCAGCGCG 2638
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
911 .SerLeuLeuGluGlyTyrAsnAlaAlaPheAsnGlyAlaIleThrGly 927
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2639 GCAAG...GATACGGCATTAACCTTAAAGACAGCGAATGAGCGTCCG 2685
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
927 lymArgAlaAspValSerMetHis.....AsnAlaLeuThrThrLeuGly 941
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2686 TCGGGCAAGCAATTAAGCAATTAACCTTGACACGCCACCATTAACACT 2735
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
942 GlyAspSerAlaIleHisThrLeuThrValArgAsnSerArgIleSer.. 957
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2736 CAATTCGGCTATCGACACGATGGGAGCGGCAACCGGCAAGTGGCG 2785
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
957 ..... 957
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2786 CAGATGGCGCGCGCGCGCTTCGCGCGCTTCCTATTATTCGTTACGCG 2835
      ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

```

957 ..... 957
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2836 CCAACTTCGGCAGATATCCGTTTCAACACGCTACGGTAAACGCAAAAT 2885
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
958 ....SerGluGlyAspArgThrPheArgThrLeuThrValAsn...LysIle 972
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2886 GAACCGTCAGGGAACA...TTCGCTTTATGTGGAATCTTCGCGTACC 2932
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
972 uAspAlaThrGlySerAspPheValIleuArgThrAspLeu.....LysA 987
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2933 GCAAGCGCAATTAAGCTGCGGAAGTTCGCAAGCACTTAACACTTG 2982
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
987 snAlaAspIysIleAsnValThrGluLysAlaThrGlySer..... 1000
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
2983 GCTGTCAACAATACCGGCAACGAAACCGGTAAGTCTGACCAATTCAGGT 3032
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1000 ..... 1000
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3033 AGTGAAGAAAGAACACACACCGCTGTCGCAAAATCTTAATTCACCC 3082
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1001 .....AspAsn.....SerLeuAsnValSerPhe.... 1008
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3083 TGCAAAACGACACGTGATGCGCGCATGGCGTTATCAGCTTATCCGC 3132
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1008 ..... 1008
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3133 AAAGACGGGAGTTCCGCTGCATTAATCCGTTCAAGAAACAGAGCTTTC 3182
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1009 .....MetLysAspProAlaGlnGlyIleSerLeuAs 1019
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3183 CGCAAACTCGGCAAGGCGGAGAAACAGACAGCGCGCTGACGGCAAAAC 3232
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1019 nIleProLeuValThrAla.....ProAlaGlyThrSerAlaGlu 1033
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3233 AGGCACAACTTGCGCGCAACAGCGGCAAAAGACACGCGCAAGC 3282
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1033 ePheLysAlaGlyThrArg..... 1039
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3283 CTGACAGCGCTGATTCGCGCGCGGC...AATGCCACGAAAGGACAGA 3329
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1040 .....MetIleGlyPheSerArgValThrProThrLeuHisValAs 1053
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3330 AAGTGTTCGCGAAACCGCGCGGACAGCGCGGGAATTCGCGGCACTTA 3379
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1053 pThr.....SerGlyLysAsnThrIleStrIleI 1063
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3380 TGCAGCGGAGAGAGAAAAAGGCTGACGCGGATTAAGACACCGCC 3429
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1063 euAspGly.....PheLysAlaGluLysAspLysAla 1073
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3430 TTGGCGAAACAGCGGCAACGCGGACCGCGGCTACACCGCTTCC 3479
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1074 AlaAlaAlaLysAlaLysPse..... 1080
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3480 CCGCGCGCGCGCGCGCGGATTTGCGCAACGCGAGCCCAACGCC 3529
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1080 ..... 1080
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3530 AACCCCAACCGCAGCGCGCACTGATCAGCGCTTAATGCCAATAGCGTTG 3579
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1081 .....PheMetAsnAlaGlyTyr 1086
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3580 AGTGAATTTTCGCGACGCTCAACAGCGTTTCGCGGTACAGAGAGATTT 3629
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1087 LysAsnPheMetThrGluValAsnAsnLeu..... 1096
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3630 GGAACCGGTGTTGCCGAAGACCGCGCAACGCGCTTGGACAAAGCGCA 3679
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1097 .....MetGlyAspL 1103
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
3680 TCCGGGACACCAAAACACTACGTTGCAAGATTTCCGCGCTTACCGGCA 3729
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
1103 euArgAspThr.....AsnGlyAspAlaGlyAlaIleArg 1115
      ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

```

3730 .....CAACCGACCTGGCCCAATCGGTATGAGAAAAACCTCG 3770
1116 ILeMeSerIglValIglSerAlaAspIglYlYrSerAspSntYrTh 1132
3771 CAGCGGCGCGGTGGCATCTGTTTTCGACACCGGACCGGA..... 3813
1132 rHISValGlnValIglPheAspIglYrSntIglLeuAspIglValAsp 1149
3814 .....AACACCTTCGACGCGCATCGCACTCGGCA 3846
1149 eupheThrIglValIglrMetThrYrThrAsp.SerSerAlaAspSerH 1165
3847 CCGGTTCCCAACGCGTCCGTCATACGCGCATCGGCGGTTTGA 3896
1165 SAIA.....PheSerIglYrThIglYrSerValIglY..GI 1176
3897 CATCGCATACGCGCGGCGGTTCAGTACGCGCATCGGCTTCAGACG 3946
1176 Y..GlyLeuYrAlaSerAlaLeuPheGlnSerGlyAlaIglYrIleAsp 1192
3947 GCATCAGACGCAAAATCCGCGCGCGCTGCATTACGCGCATTCAGCA 3996
1192 euIle...GlyLys.....TyrIleHisHis.....AspAsn 1201
3997 AGATACCGCGGAGTTTCGGCGGATTCGGCATGCAACCGCATCGCGC 4046
1202 AspYrThrIglYrAlaPheAlaIglYrLeuGly..... 1211
4047 AACCGGTATTCGTCCCAAAAGCGGATTCACGATACGAAAGTCATA 4096
1212 .ThrIleHisIglYrAlaSntIglYr..... 1221
4097 TCGCCACCGCGCGCTTCGATTACACCGCTACCGCGCGCATTAAGCA 4146
1222 .....AlaGlyAlaGluThr 1226
4147 GATTATTCATTCAAACCGCGCACACATTCATTCACGCTTATTCAG 4196
1227 GlyYrThrIglYrAlaSntIglYrPheIleGluProGlnAlaGI 1243
4197 CCGTCTCTATACGCGTCCGCTTCGCGCAA..... 4227
1243 uIleuValYr...GlyAlaValSerGlyIglYrPheArgTrpIglYr 1259
4228 .....GTCCGACGCGCGCTCATACCGCGCTTTCGCG 4260
1259 IYAspMetAlaPleuSerMetLysAsnArgAspPheSerProIleuIleGly 1275
4261 CAG.....GATTTCGCGCAAAACCGCGCATGCG...GAATGGGCGT 4298
1276 ArgThrIglYrIleGlnLeuGlyIglYrPheSerIglYrAspTrpSerVa 1292
4299 AAACGCC 4305
1292 IThrAla 1294
seq_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA197.DAT:AAW27705
seq_documentation_block:
ID AAW27705 standard; Protein: 323 AA.
XX
XX AAW27705;
XX
XX 08-MAY-1998 (first entry)
XX
XX H. influenzae Hap protein autotransporter membrane integration region.
XX
XX Hap protein; autotransporter; Gram-negative bacteria; diagnostic;
XX
XX therapy; surface presented polypeptide.
XX
XX Haemophilus influenzae.
XX

```

```

PN WO9735022-A1.
XX
XX 25-SEP-1997.
XX
XX 15-MAR-1996; 96WO-EP01130.
XX
XX 15-MAR-1996; 96WO-EP01130.
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Jose J, Maurer J, Meyer TF;
XX
XX WPI: 1997-480227/44.
XX
XX N-PSDB: AAT88142.
XX
XX Presentation of peptide(s) on surface of Gram-negative bacteria -
XX via transformation with vector encoding signal peptide, presented
XX peptide and transporter domain of auto-transporter, producing
XX peptide libraries for epitope mapping
XX
XX Claim 8; Fig 9; 84pp; German.
XX
XX This sequence represents the H. influenzae Hap autotransporter membrane
XX integration region. This region is involved in a novel method which
XX allows the presentation of stable fusion polypeptides on the surface of
XX Gram-negative bacteria which can be released into the surrounding media.
XX The method can be used to produce a variegated population of
XX surface-presented polypeptides, so that bacteria expressing polypeptides
XX with particular properties can be identified and simultaneously selected,
XX e.g. for epitope mapping or selection of ligands with the highest
XX affinity for antibodies, major histocompatibility complex (MHC) molecules
XX or other components of the immune system. Selected polypeptides can be
XX used diagnostically, e.g. to screen sera or antibody banks, and
XX polypeptide expressing cells may be used as live vaccines. They may also
XX be used therapeutically, e.g. when the polypeptide is an antibody, to
XX remove or concentrate pollutants, inactivate toxins, prepare and process
XX food, prepare washing compositions and label cells. Selected bacteria can
XX be stored, reproduced and replicated on a large scale as individual
XX clones.
XX
XX Sequence 323 AA;
XX
XX alignment_scores:
XX Quality: 721.50 Length: 308
XX Ratio: 2.898 Gaps: 3
XX Percent Similarity: 80.844 Percent Identity: 45.130
XX
XX alignment_block:
XX US-09-303-518d-653 x AAW27705 ..
XX
XX Align seg 1/1 to: AAW27705 from: 1 to: 323
XX
XX 3487 CGCGCGCGCGCGCGGAT...TTGCGCAACCGCGACCGCCGCAACC 3533
XX .....
XX 17 GlnGlnSerGlnIglYrAspArgLeuAlaGlnIglYrAlaGlnIglYrSln.. 32
XX
XX 3534 CCAACCGCAGCGCGACCTGATCAGCGCTTATGCCATACGCGGTTCAGTG 3583
XX .....
XX 33 .ArgIglSlnIglYrAspLeuIleSerArgYrSerAsnSerAlaLeuSerG 49
XX
XX 3584 AATTTTCGCGCGCTCAACAGCGTTTCGCGCTACAGACAAATTTGAGC 3633
XX .....
XX 49 IuLeuSerAlaThrValAsnSerMetLeuSerValGlnIglYrIleuAsp 65
XX
XX 3634 CCGGTGTTTCGCGAGAGACCGCGCAACGCGCTTTCGCAACGCGCATCGC 3683
XX .....
XX 66 ArgLeuPheValAspIglAlaGlnSerAlaValIglYrThrAsnIleAlaGI 82
XX
XX 3684 GGAACCAACCACTACCGCTTCGCGCAAGATTCGCGCGCTACCGCAACAA. 3732
XX .....
XX 82 nasplysArgArgYrAspSerAspAlaPheArgAlaYrGlnGlnIgl 99
XX

```

[illegible]

```
seq_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:AAp80136
seq_documentation_block:
```

seq_documentation block: ID	AA#0136 standard; protein; 741 AA.
XX	AA#0136;
AC	
XX	
DT	09-OCT-1990 (first entry)
XX	
DE	Neisseria IGA-Protease precursor protein
XX	
KW	IGA-Protease precursor; Neisseria sp;
KW	Gram-negative bacterial live vaccines.
XX	
OS	Neisseria gonorrhoeae (strain SMH).
XX	
PH	
Key	Location/Qualifiers

FT	Cleavage-site	193..197
FT	/label-IgA-Protease cleavage site (a)	
FT	Cleavage-site	226..229
FT	/label-IgA-Protease cleavage site (b)	
FT	Cleavage-site	327..332
FT	/label-IgA-Protease cleavage site (c)	
XX		
PN	DE3622221-A.	
XX		
PD	14-JAN-1988.	
XX		
PF	02-JUL-1986;	86DE-36622221.
XX		
PR	02-JUL-1986;	86DE-36622221.
XX		
PA	(PLAC ) MAX PLANCK GES WISSENSCH.	
PI	Meyer TF, Halter R, Pohlner J;	
DR	WPI; 1988-015104/03.	
DR	N-PSDB; AAN60154.	
XX		
PT	Extracellular prodn. of proteins -	
PT	by gram-negative host cells contg. a vector conty. one or more	
PT	genes coding the desired protein	
XX		
PS	Disclosure; ; German.	
XX		
CC	Precursor protein consists of three regions i.e. Amino terminal	
CC	leader sequence, IGA protease and a "helper" domain. The cleavage	
CC	sites given in the features lie in the region between the latter	
CC	two domains. DNA encoding a desired protein can be cloned into the	
CC	corresponding region in the IGA protease precursor gene, between	
CC	the DNA that encodes the natural cleavage sites. Thus, IGA protease	
CC	coding region is not disrupted and the desired protein is released	
CC	following cleavage by the protease.	
XX		
SO	Sequence	741 AA;

alignment_scores:	
Quality:	442.50
Ratio:	1.114
Gaps:	19
Percent Similarity:	47.903
	Percent Identity: 22.490
alignment_block:	

Align seg 1/1 to: AAP80136 from: 1 to: 741

2521 CTTTCGACAAAGCTAAGGCAAAAGCTAAGCCATTCCGCACTCAACGGCAA 25700  
|||||||:::| |:::| |:::| |:::| |  
1 LeuSerAspIysAlaLeuAsnSerPheAspAlaThrProIleAsnGlyAs 17

```

2571 TGTCCTCCCTAGCCGATAGGCAGCTATTCCATTTTGAAAACAGCCGCTTTA 2620
      |||::||| :::::||::: ::::: :::::
17 nvaIasnIeuAasnGInasnaIaIaIeuValIeuGlyLysaIaIaIaIeuT 34

```

```

2621 CCGGAATAATCAGCGCGCGGACGATACGGCATTTACACTTA...AAAGAC 26
      |||||:::|  ::::  ::  ||  :::
34  rrglyylslleglnlglnglglasnserargvalserleuasnlnl  50

```

```

2668 AGCGAATGAGACGCTGCCGTCGGGCACGGAAATTAGGCATTTAAACCTTGA 27
      |||::|||   |||   ::   :::::   |||::|::|
51 SerLysTrpHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAl 67

```

```

218 CAACGCCACCATTCACATCCGCGCTATCGACACGATGCGCGACGG 2767
      ::::: ||| |||||:::|||| ||||| :::|
67 aaspSerHisIleHisLeuasnAsnAla.....SerAspalaGlnSerA 82

```

2768 CGCAACCGGAGTGGGAGATGGCCGGCCGCTTGGCGCGTTC 2817  
 ||  
 82 la..... 82

[illegible]

```

657 PheAlaTyrArgPheLys...ValGlyAsnLeuThrValLysProLeuLe 672
4194 GAGCGCTGCTATACCGATGCCGCTCCGGAAGTCCGAACGGCGCTCA 4243
672 uSerAlaAlaTyrPhe...AlaAsnTyrGlyLysGlyAlaValAla 688
4244 ATACCGCGGTATGCGCAGGATTTCCGCAAAACCCGACAGCGGAAATGG 4293
688 LysGlyLysSerPheAlaTyrLysAlaAspAsnGlnGlnTyrSerAla 704
4294 GCGGTAAACGCCGAATCAAGTTTCAACGCTGTCCTCCACGCTGCCGC 4343
705 GlyValAlaLeuLeuTyrArgAsnValThrLeuAsnValAlaSerIle 721
4344 CGCCAGAGGCGCGCAATTCGAGAGCGACAGCGCGGCGGATCAATTAG 4393
721 eThrLysGlyLysGlnLeuGlnLysGlnLysSerGlyGlnLysIleG 738
4394 GCTACCGCTGG 4404
738 InileArgPhe 741

seq_name: /STDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA98842
seq_documentation_block:
ID   AA98842 standard; Protein: 1569 AA.
AC   AAG98842;
XX
DT   26-SEP-2001 (first entry)
XX
DE   E. coli growth and proliferation related protein sequence SEQ ID NO:312.
XX
KM   Escherichia coli; growth; proliferation; microbial; antimicrobial;
XX   bacterial infection; microorganism.
XX
OS   Escherichia coli.
XX
PN   WO200134810-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30950.
XX
PR   09-NOV-1999; 99US-0164415.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA, Ohlsen K, Zyskind J;
XX
DR   WPI; 2001-335933/35.
XX
DR   N-PSDB; AAH84513.
XX
PT   Novel nucleic acids that inhibit Escherichia coli proliferation, useful
XX   for screening for homologous genes and for designing expression vectors
XX
PS   Claim 19; Page 396-399; 522pp; English.
XX
CC   AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
XX   related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
XX   growth and proliferation related proteins given in AAG99078 and AAG98830
XX   to AAG98999. (1) can be used as potential targets for the generation of
XX   new antimicrobial agents, and for identification of compounds which
XX   interact with the gene products of (1). In addition the expression of
XX   (1) and the purification of the proteins, the purified proteins can be
XX   used to generate reagents and screen small molecule libraries or other
XX   candidate compound libraries for compounds that can be further developed
XX   to yield novel antimicrobial compounds. In addition, nucleic acid probes
XX   complementary to (1) that are specific for particular species of
XX   microorganisms can be used to identify particular microorganism species
XX   in clinical specimens, therefore, providing a rapid and dependable

```

```

CC   method by which to identify the causative agents of a bacterial
CC   infection. Also, antibodies generated against proteins translated from
CC   mRNA transcribed from proliferation-related sequences can also be used
CC   to screen for specific microorganisms that produce such proteins in a
CC   species-specific manner. AAH84371 and AAH84670 represent sequencing
CC   primers used in the isolation of E. coli growth and proliferation
CC   related sequence, which are used in an example from the present
CC   invention.
XX
SQ   Sequence 1569 AA;

alignment_scores:
    Quality: 359.00      Length: 1686
    Ratio: 0.481        Gaps: 80
    Percent Similarity: 44.306    Percent Identity: 10.395

alignment_block:
US-09-303-518D-653 x AAG98842 ..

Align seg 1/1 to: AAG98842 from: 1 to: 1569

445 GGGACTAACGGCCATCTTATGCGCGGATATCATATGCGCGCTTTCGA 494
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 GlyThrThrAsnAsnThrHisIleIleAsnHisGlyGlnAsnValHis 113
495 CAATTTTCACAGATGACGAACTTTCAGATGACCACTTATATGATG 544
|-|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 sGlyGlyValSerAsnGlySerLeuIleGlySerGlyGlyTyrGlnAspI 130
545 GGTGAATATACGCTGATTAATTAATACCTGATGCTGTGATGCGGA 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 IeGlySerHisAsnAsnPheValGlyGlnAlaAsnHisThrThrIleAsn 146
595 GCAGCGACGACAAATATTGGCGGTGTGATGAAGACGAAACCAATACCGCGA 644
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 GlyGlyArgGln..... 150
645 AAGTTCATATCATATTGCA.....AGCGCATATCTTGCGCTCGGTG 688
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 ...SerIleHisAspGlyGlyIleSerThrGlyThrThrIleGlnSerG 166
689 GCAATACCTTTGCACAAATGGA.....TCAGGTGGTGCACAGTCAC 732
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 LysAsnGlnAspValTyrLysGlyGlyIleSerAsnGlyThrThrIleLys 182
733 TTAGGTAGCGGAAAAATTAA...CATAGCCCATATGCTTTTATCAAC 779
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 GlyGlyAlaSerArgValGlnGlySerAlaAsnGlyIleLeuIleAs 199
780 AGGAGGCTCA.....T 790
199 polyGlySerGlnIleValLysValGlnGlyHisAlaAspGlyThrThrI 216
791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGACCAAAAGCAAA 840
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 LeAsnLysSerGlySerGlnAspValValGlnGlySerLeuAlaThrAsn 232
841 TCGTTATTAATGCGGTATTCGCAACAGCAACCCCTATATAGA..... 885
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 ThrThrIleAsnGly.....GlyArgGlnTyrValGlnGlnIle 245
886 .....AAAGCAATGGCTTCACAGCTAGTTCGTA 913
245 rThrValGlnThrThrThrThrIleLysAsnGlyGlyGlnGlnAspValTyrG 262
914 AAGATTGCTTCATGATGAATCTTTGCTGAGATACCATTCAGTATTC 963
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
262 LysSerArgAlaLeuAspThrThrThrIleGlnGlyGlyThrGlnSerLeu... 277
964 TACGAAACCAATCAAAATGGGAAATACTTTTAAACGCAATATATATG 1013
277 ..... 277

```

```

1014 CGCAGAAATGATGCCAAACATAACACTATCTACTATAGAT 1063
278 .AsnSerLysThrAlaLysAsnThrHisIle.....TyrSerG 291
1064 TAAAAACACCAACCGTCAATGTTAAATGTTCTTATCCGAGACAGA 1113
291 LysGlyThrGlnIleValAsp.....AsnThrSerThrSerAspValIle 305
1114 AGAGAACCTTTATCATGCTGCAGGTGGGAGTCAACAGTTATGACCCAG 1163
306 Gln.....ValTyr.....SerGlyGlyValLeuAspValArg..... 316
1164 ACTGATATATGAGAAATATTTCCCTTATGACAAAGAAAGTGAAAT 1213
317 .....GlyGlyThrAlaThrAsnValThrGlnHisAspGlyAlaI 330
1214 TGATACTTACACGACATCAACCAAGCGCGGCGTTGTATTTGAG 1263
330 LeuLysThrAsnThrAsnGlyThrThrValSerGlyThrAsnSerGlu 346
1264 GGTATTTTACGGCTCTGCCCTAAACACAGAAACCGTGCAGAGCGCGG 1313
347 GlyAlaPheSerIle.....HisAsnHisVal.....AlaAspAs 358
1314 CGTTCATATCATGATGTCAGTACCTTACTTGTGAAGTAAAGCGCGTGC 1363
358 nValLeuLeuGlnAsnGlyGlyHisLeuAspIleAsnAlaTyrGlySerA 375
1364 CAACAGCACCGCGCTGCCAAATCGGCAAGGCGACGCTG.....CTGGTT 1407
375 LaaSnLysThrIleIleLys...AspLysGlyThrMetSerValLeuThr 390
1408 CAACGCCAAGGG.....GAAACCAAGCGTCCGCGACGCT 1442
391 AsnAlaLysAlaAspAlaThrArgIleAspAsnGlyGlyValMetAspVa 407
1443 GGGGAGCGGT.....AAAGTCATCTTATGATCAGACGAGCGACGATCAAG 1486
407 LAlaGlyAsnAlaThrAsnThrIleIleAsn.....G 418
1487 GCAGAAACCAAGCCTTATAGTGAATCGGCTTGGTCACCGGC..... 1527
418 LysGlyThrGlnAsnIleAsnAsnTyrGlyIleAlaThrGlyThrAsnIle 434
1528 ...AGGGGAGCGGTGTCACATGATCCGATATCATGTTCAACCCGCAAA 1574
435 AsnSerGlyThrGlnAsnIleLysSerGlyGlyLysAlaAspThrThrI 451
1575 ACTGTATTTGCGCTTTCGC.....G 1594
451 eIleSerSerGlySerArgIleValIleGluLysAspGlyThrAlaIleG 468
1595 GCGGACGTTTGATTTGAACGCGCATCGCTTTCGTC..... 1632
468 LysSerAsnIleSerAlaGlySerLeuIleValIlyThrGlyGlyIle 484
1633 ...CACCGCATTCAGAAATACCGATGAAGGCGCATGATTGTCAC..... 1674
485 AlaHisGlyValAsnGlnGluThrGlySerAlaLeuValAlaAsnThrG 501
1675 .....CACATCAAGACAAAGAAATCCACCGTTACCATTA 1708
501 yAlaGlyThrAspIleGlnGlyTyrAsnLysLeuSerHisPheThrIle 518
1709 CAGGCATTAAT.....GATATACT 1728
518 hrgGlyGlyGluAlaAsnTyrValValLeuGlnAsnThrGlyGlyLeuThr 534
1729 .....ACACCGGCATTAACACAACTTGATAGCAAAAGA 1766
535 ValValAlaLysThrSerAlaLysAsnThrThrIleAspThrGlyGlyL 551
1767 AATTGCCTCAACAGCGTTGTTGGCAGACAAAGATGCAACCAACAGACG 1816
551 sLeuIleVal.....GlnLysGluAlaLysThrAspSer 563
1817 GCGGCTCAAT.....CTG 1830
563 hArgLeuAsnAsnGlyGlyValLeuGlnValGlnAspGlyGlyGluAla 579
1831 AATTACCAACCGAAGACGATCGACTTACGCTTTC..... 1872
580 LysHisValGlnGlnGlnSerGlyGlyAlaLeuIleAlaSerThrThrSe 596
1873 .....GCGGACAAATTTAACGGCAAT..... 1896
596 rGlyThrLeuIleGlnGlyThrAsnSerTyrGlyAspAlaPheTyrIleA 613
1896 ..... 1896
613 rGAsnSerGluAlaLysAsnValValLeuGlnAsnAlaGlySerLeuThr 629
1897 .....ATCACGCAACAAACGCGCA 1916
630 ValValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGlyL 646
1917 ACTGTTTTCAGCGCAGACGACACCGCGCTTACATCATTTAGCAA 1966
646 sMetAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyT 663
1967 GC..... 1968
663 hrgGlnThrIleTyrAlaSerAlaThrSerAspLysAlaAsnIleLysGly 679
1969 .....GCGGTGCAAAATGAGAGGTATCCCAACG 2000
680 GlyLysGlnThrValTyrGlyLeuAlaThrGluAlaAsnIleGlnSerG 696
2001 AGAAATCGGTGCGACACAGATGATGATGACCGCATTTAAAGCGAAA 2050
696 yGlnGlnIleValAlaAspGlyLysThrThrGluLysThr..... 708
2051 ACTTCATATTCAGGCGCA..... 2070
709 .....HisIleAsnGlyGlyThrGlnThrValGlnAsnTyrGlyLysAla 723
2071 .....CAGCGGTGTTTCCCGCAATGT 2093
724 IleAsnThrAspIleValSerGlyLeuGlnGlnIleMetAlaAsnGlyTh 740
2094 TCCCAAA.....GTGCAAGCGGATGCGCATTAAGCATCAACGCC 2134
740 rAlaGlnGlySerIleIleAsnGlyLysSerGlnValValAsnGlnGlyG 757
2135 AAGCATTTTGGTGTGCGACCGCATCAAGCCACACATCTGTACAGT 2184
757 LysLeuAlaGlnAsnSerValLeuAsnAspGlyGlyThrLeuAspValArg 773
2185 .....TCGAGCTGACGCGCTCG..... 2202
774 GluLysGlySerAlaThrGlyIleGlnGlnSerGlnGlyAlaLeuVa 790
2203 .....ACAAGTTGACGCAAAACCAATTACCGACGATTAAGTATG 2245
790 lAlaThrThrArgAlaThrArgValThrGlyThrArgAlaAspGlyValA 807
807 lAheSerIleGlnGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsn... 822
2246 CTTCATTTAGCAAGACGACATCAGACGCAATGTCAGCTTCGCGATCAC 2295
822 lAheSerIleGlnGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsn... 822
2296 GCTCATTTAAATCTCACAGCACTGCGACACTCAACGCAATCTTAGTGC 2345
823 .....GlyGlyValLeuThrValGlnSerAspThrSerSe 834
2346 A.....GCGGACACACAGCACTTACGGTTACGC 2374

```



```

834 rAspIysThrGluValAsnMetGlyGly.....ArgGluIleValIleYST 849
2375 GCAAGCGCCGCAAAAGCGCAACCTCAGCGCTGGGCAATGCCCA... 2421
      ::::::::::::::::::::
849 hrlYsAlaThrAlaThrGlyThr.....ThrLeuThrGlyGlyGluIle 864
2422 .....GCAACATTATATCAAGCCACATTAAAC.....GGCAACACATC 2459
      ::::::::::::::::::::
865 ValGluGlyValAlaAsnGluThrThrIleAsnAspGlyGlyIleGluThr 881
2460 GGCTTCGGCAATGCTTCTCATTTAACTAAGACACAGCCGTACAAACG 2509
      ::::::::::::::::::::
881 rValSerAlaAsnGlyGlu..AlaIleIleThrIleAsnGlyGly 897
2510 GCAGTGTGACGCTTCCGACAAAGCTAAGGCAAC.....GTAGGCAT 2553
      ::::::::::::::::::::
897 lYThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValGluIleAsn 913
2554 TCAGCACTCAACGCGCAATGTCTCCCTAGCCGATTAAGCAGTATTCATT 2603
      ::::::::::::::::::::
914 SerGlyAlaAlaIleuGluIleThrSerThrAlaAsnGlyIleGluIleSerG 930
2604 TGAACACGCGCCCTTACCGGAAATACAGCGCGCAGAGATACGCAT 2653
      ::::::::::::::::::::
930 YThrIleGluIleThrPheSerIleSerGlyAsnLeuAlaThrAsnM 947
2654 TACACTTAAGACAGC..... 2670
      ::::::::::::::::::::
947 eLleuLeuGluAsnGlyLysLeuLeuValIleAlaGlyThrGluAla 963
2671 ...GATGGACGCTGCGCGGCAAGCAATTAAGCAATTAACCTTGA 2717
      ::::::::::::::::::::
964 ArgAspSerThrValGlyGlyAlaMetGlnAsnLeuGlyGlnAs 980
2718 CAAGCGCCACATTACACTCAATTCGCCCTATCGACAC..... 2754
      ::::::::::::::::::::
980 pSerAlaThr...LysValAsnSerGlyGlyIleThrLeuGlyArgS 996
2755 .....GATCGGCA 2763
      ::::::::::::::::::::
996 eLrYsAspGluPheGluAlaLeuAlaArgAlaGluAspLeuGluValAla 1012
2764 GGC.....GCGCAACCGCGCAGTGGCGAGATGGCGCGCGC 2801
      ::::::::::::::::::::
1013 GlyIleThrAlaIleValIleValIleThrLeuAlaAspAlaSerVal 1029
2802 CCGTTCGCGCGCTCCCTATATCCGTTAGCGG..... 2835
      ::::::::::::::::::::
1029 rGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnValThrp 1046
2835 ..... 2835
      ::::::::::::::::::::
1046 roValLysLeuGluGlyAlaValAlaArgIleThrAspSerAlaThrLeuThr 1062
2835 ..... 2835
      ::::::::::::::::::::
1063 LeuGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSerAr 1079
2835 ..... 2835
      ::::::::::::::::::::
1079 gGlySerValThrLeuAsnSerAsnAsnSerGlyAlaGlyThrSerAsn 1096
2835 ..... 2835
      ::::::::::::::::::::
1096 ysgIuYrArgValAsnSerLeuLeuAsnAspGlyAspValTYrLeu 1112
2836 .....CCAAGTGGCAGAAATCCCGTTTCAACAGCT 2867
      ::::::::::::::::::::
1113 SerIleAlaThrAlaAlaProAlaThrThrAsnGlyIleTYrAsnThrIle 1129
2868 GAGCGTAAAGCGCAATTTGAACGCTCAGGACATTCGCTTATGTCGG 2917
      ::::::::::::::::::::

```

```

1129 uThrThrAsn...GluLeuSerGlySerGlyAsnPheTYrLeuHisThrA 1145
2918 AACTCTTCGCTACCGCAGCGGCAATTAAGAGCTGGCGGAATTCGAA 2967
      ::::::::::::::::::::
1145 snValAlaGlySerArgGlyAspGluLeuValValAsnAsnAlaThr 1161
2968 GGCATTACACACCTTGCTGTCACCAATACCGGACAGCAACCCGTAAGTCT 3017
      ::::::::::::::::::::
1162 GlysnPheLysIlePheValGlnAspThrGlyValSerProGlnSerAs 1178
3018 CGAGCAATTGACGGTAGTGAA.....GGAAAAACACACACCGCTGT 3061
      ::::::::::::::::::::
1178 pAspAlaMetThrLeuValLysThrGlyGlyAsp..... 1190
3062 CCGAAATCTTAATTCACCTGCAAAAGCA.....CAGTCGATGCC 3105
      ::::::::::::::::::::
1191 .....AlaSerPheThrLeuGlyAsnThrGlyGlyPheValAspLeu 1204
3106 GCGCATGCGGCTTATCAGCTTATCCGCAAGAGCGGAGTTCGCGCTGCA 3155
      ::::::::::::::::::::
1205 GlyThrTYrGluTYr...ValLeuLysSerAspGly..AsnSer..... 1217
3156 TAATCCGTCGCAAGAACAGAGCTTCCGACAAACTCGCAAGCGGAG 3205
      ::::::::::::::::::::
1217 ..... 1217
3206 AACACAGAGCGCGCTTGACGCAAAACAGCACACTTCGCGCAACAA 3255
      ::::::::::::::::::::
1217 ..... 1217
3256 CAGCGGAAAAAGACAAAGCGCAAGCTTGACGCGTATTCGCGCGG 3305
      ::::::::::::::::::::
1218 .....AsnThrAsnLeuThrAsnAspValLysProA 1228
3306 GCCCAATGCCACCGAAAAAGCAGAAAGTGTGCCGAGCGCGCGGACG 3355
      ::::::::::::::::::::
1228 snProAspPro.....IleProAsn..... 1234
3356 CAGCGGGGAAATGCCGCGATTATGCAGCGGAGGAGCAAGAAAAACGG 3405
      ::::::::::::::::::::
1234 ..... 1234
3406 GTCCAGCGGATTAAGACACCGCTTGCGGAAACAGCGGAAGCGGAAC 3455
      ::::::::::::::::::::
1235 .....ProLysP 1237
3456 CCGCGCGGCTACCAACCGCTTCCCGCGCGCGCGCGCGCGGAT 3505
      ::::::::::::::::::::
1237 roAspProLysProAspProLysProAspPro..AsnProLysProAsp 1252
3506 TGCCGCAACCGGAGCCCAAGCGCAACCGCAACCGGAGCGGACCTGATC 3555
      ::::::::::::::::::::
1253 ..ProThrProAspProThrProThrProValProGluLysArgIleThr 1268
3556 AGCGGTTATGCCAATAGCGGTTGAGTGAATTTTCGCCACCGCTCAACAG 3605
      ::::::::::::::::::::
1269 Pro.....SerThrAlaAlaValLeuAsnMetAlaIleThrLeuProIle 1283
3606 CGTTTC.....GCCGTACAGACGAATTTGACCCCGCT 3640
      ::::::::::::::::::::
1283 uValPheAspAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetL 1300
3641 TTGCCGAAAGACCGCGCAACGCGGTTTGCAAGGAGTTCGCGGACAC 3690
      ::::::::::::::::::::
1300 yAlaSerProHisAsnAsnAsnValThrp...GlyAlaThrTYrAsnThr 1315
3691 AAACACTACCGTTGCGCAAGATTTCCGCGCTACCGCAACAAACGAGACT 3740
      ::::::::::::::::::::
1316 ArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGluGluThr..... 1330
3741 GCGCCAAATCGGATATGCAAAAAACCTCGCGCAGCGGCGC.....G 3781
      ::::::::::::::::::::
1331 ....LeuThrGlyMetThrValGlyIleAspSerArgAsnAspIleProG 1346

```

```

3782 TCGGATCTCTG.....TTTTCGCAACGCGGAC 3810
      |||||
1346 lucyllethrhleuglyalaphemetyltyrserhisserhisile 1362
      |||||
3811 GGAACACCTTCGACGAGCATCGGCACTCGCACGGCTTCCGACGG 3860
      |||||
1363 gly.....Pheaspargly.....Glyhisgl 1370
      |||||
3861 TGCCGTTTGGGCAATAGCGCATCGC..... 3888
      |||||
1370 yserVal...GlyserTyrserleuglyglytyrAlaserTrpLuhisg 1386
      |||||
3889 .....ACGTTCGACATC 3900
      |||||
1386 luserglyPheTyrleuaspolyalValylsleuasnAtrpPheylsSer 1402
      |||||
3901 GGCATCAGC.....GCGGGCGCGGTTTATAGCGGACGCTT.. 3939
      |||||
1403 AsnValAlaaglysmetserserglyalAlaAlasnGlyserTyrH 1419
      |||||
3940 .TCAGAGCGCATCGAGGCAAAATCCCGCGCGCTCTGCATTACGGCA 3988
      |||||
1419 sSerasnGlyleuglyglyHis.....AGTTTCGATTCACCGCT 1427
      |||||
3989 TTGAGCAAGATACCGCGGAGTTTCGGGATTCGCGATCGAACCGCAC 4038
      |||||
1427 legluthrGlymetAtrpPheThrAspGlyasnTrpAsnleuThrProTyr 1443
      |||||
4039 ATCGGCGGACGCGCTATTTCTGTCGCAAAAGGAGATTACCGATCGAA 4088
      |||||
1444 AlaserleuThrGlyPheThrAlaAspAsnProGlyTrhisleuSerAs 1460
      |||||
4089 C.....GTCAATATCGCCACCGCGCGCTTGACATTCACCGCT 4126
      |||||
1460 nGlyMetlysserlysserValAspTrpHrArgserile.....T 1473
      |||||
4127 ACCGC...GCGGCGATTAAAGGAGATTATTCATTAACGCGGCAACAC 4173
      |||||
1473 yTrArgGlyleuglyAlaThrleuSerlyrAsnMetAtrgLeuGlyAsnGly 1489
      |||||
4174 ATTTCCATCAGCGCTTATTTACGCTTCCATATCCGATCGCGCTTCGG 4223
      |||||
1490 MetGlyValAlaGlyProTrpLeuLysAlaAlaValAtrgylsleupheValas 1506
      |||||
4224 CAAGTCCGACGCGCGCTACATACCGCGGTATTCGCGGAGGATTC... 4269
      |||||
1506 pAspAsnAtrgValylsValAsnSerAspGlyAsnPhelValAsnTrpLeus 1523
      |||||
4270 .GGCAAAACCCGAGTCCGGAATGGGCGTAAACGCGCAATCAAAAGT 4317
      |||||
1523 erglyAtrgAtrgGlylYletrglnAlaGlylYlelYsAlaserPheSer 1539
      |||||
4318 TTCACGCTGCTCCCTCCACGCTCCGCGCCCAAGGCGCGCATATGGAAC 4367
      |||||
1540 ...ThrleuSerGlyHisleuGlyValGly.....Ty 1549
      |||||
4368 GCAGACAGCGGCGGCGGATCAAA.....TTAGGCTAAC 4399
      |||||
1549 rSerHisserAlaGlyValGlyUserProTrpAsnAlaValAlaGlyVala 1566
      |||||
4400 GCTGG 4404
      |||||
1566 snTrp 1567
      |||||
seq_name: /SIDSL/gcgdata/geneseg/genesegp-emb1/AA2000.DAT: AAB01830
seq_documentation_block:
ID AAB01830 standard; Protein; 1222 AA.
XX
AC AAB01830;
XX
DT 11-SEP-2000 (first entry)

```

```

XX
DE H. Influenzae strain K1 mature full-length HmWA protein, SEQ ID NO:37.
XX
KW Mature HmW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typable Haemophilus influenzae; NTH1; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain K1.
XX
FH Key Location/Qualifiers
FT Misc-difference 307
FT /note= "Encoded by GG"
XX
PN WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CM00938.
XX
PR 07-OCT-1998; 98US-0167568.
XX
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI: 2000-303789/26.
XX
DR N-PSDB: ABA52180.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 20A-R; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typable (non-encapsulated) H.
CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HmWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HmWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HmWA. The invention also discloses hmwA genes (ABA52175-A52198)
CC and HmWA proteins (AAB01824-B01849) from the non-typable H. influenzae
CC strains JcyC, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to protect a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HmWA protein from a non-typable
CC strain of H. influenzae.
XX
SQ Sequence 1222 AA;

```

```

alignment_scores:
  Quality: 310.50      Length: 1304
  Ratio: 0.475        Gaps: 62
Percent Similarity: 50.153  Percent Identity: 19.785
alignment_block:

```

US-09-303-518D-653 x AAB01830 ..

Align seg 1/1 to: AAB01830 from: 1 to: 1222

```

502 GTCACAGATGCAAGACCTGTTGAGATGACCAAGTATATATGATGGGTGAA 551
      ::::::::::::::::::::
95  IlerhrSerThrThrsnGlyAsnLeuThrIleYrSerGlyGlyTrp.. 110
      ::::::::::::::::::::
552 ATACCGTGAATTAATAATACCCGATCGCTGTCGAATCGGACGAGC. 600
      ::::::::::::::::::::
111 ....ValaspValHisLys.....AsnIleThrLeuLysSerGlyT 123
      ::::::::::::::::::::
601 .....AGACAATATTGGCGTCTGATGAGACGAACCC 633
      ::::::::::::::::::::
123 YrLeuAsnIleThrThrsGlnGlyAspIleAlaPheGluAspLysPro 139
      ::::::::::::::::::::
634 AAT.....AACCG 641
      ::::::::::::::::::::
140 GlyLeuSerAsnLeuThrIleThrAlaLysGlyThrIleAlaValAsnAs 156
      ::::::::::::::::::::
642 CGAAAGTTTCATATCATATTGCAAGCGCATATTCTGGCTCGTGGTGCA 691
      ::::::::::::::::::::
156 nLysLysGlyPheArgPhe.....A 163
      ::::::::::::::::::::
692 ATACCTTTGCACAAATGATGATGAGTGGTGACAGTCAACTTAGTAGTC 741
      ::::::::::::::::::::
163 spAsnValThrLeuAsnGlyThrGlyGlyLeu..... 174
      ::::::::::::::::::::
742 GAAAAAATTAAACATAGCCCATATGTTTATTACCAACAGGA..... 783
      ::::::::::::::::::::
175 .....SerPheLysThrIleGluThrGlyAsnArgAs 185
      ::::::::::::::::::::
784 .....GGCTCATTTGGGACAGTGGCGTCAC 808
      ::::::::::::::::::::
185 pSerAsnPheGluThrThrHisPheArgGlyArgLeuAsnIleSerGlyLysV 202
      ::::::::::::::::::::
809 CAATGTTTATCTATGATGCCCAAAAGCAAAAGTGAATTAATGAGGGTA 858
      ::::::::::::::::::::
202 alaPrlIleuMetGlnAlaArgGlnGlnLysnTrp..... 213
      ::::::::::::::::::::
859 TTGCAACAGGCAACCCCTATATAGGAAAAAGCAATGCTTCCAGCTAGT 908
      ::::::::::::::::::::
214 .....As 214
      ::::::::::::::::::::
909 TCGTAAGATTGTTCTATGATGAATCTTGTGAGATACCATTCAG 958
      ::::::::::::::::::::
214 nArgArgHisTrp.....GlyArgSerHisTrpA 224
      ::::::::::::::::::::
959 TATCTAGCAACCAATCAAAATGGGAATACCTTTTAAAC..... 999
      ::::::::::::::::::::
224 snValThrArgLeuAsnValSerGluAsnSerYrPheAsnValThrIle 240
      ::::::::::::::::::::
1000 .....GACAATTAATATGGCGGACGAGAAATGCGTGCACAA 1034
      ::::::::::::::::::::
241 AspSerSerGlySerAlaSerSerProGlyAlaGlyProLeuAsnAlaG 1257
      ::::::::::::::::::::
1035 ACATAAACAATCTTCTCTACCTTATAGATTAAACACGAACCGTTCAAT 1084
      ::::::::::::::::::::
257 nSerGlyLeuAsnGlyLysSerPheAsnAsnAsPThr.....V 270
      ::::::::::::::::::::
1085 TGTTATATGTTCTTTATTCGACAGACAGACAGAACCTGTTTATCATGCT 1134
      ::::::::::::::::::::
270 alPheAsnIleAla..... 1175
      ::::::::::::::::::::
1135 GCAAGTGGGGTCAAC...AGTTATCGACCC.....AGACGTAA 1169
      ::::::::::::::::::::
276 SerSerAlaValAsnPheAsnIleLysProIleValAspLysValTh 292
      ::::::::::::::::::::
1170 TAATGGAAGAAATTTTCCCTTTATTGACAAAGAAAGGTGAATGATAC 1219
      ::::::::::::::::::::
292 rAsnGlyAsnHisThrLeuPhe.....LysGlyAsnIleSerV 305
      ::::::::::::::::::::
1220 TTACACACATCAACCAAGGCGGGCGGCTTGTATTGTGAGGTAAAT 1269

```

```

305 alleu.....GlyGlyGlyAspValAsnPhe.....His 314
      ::::::::::::::::::::
1270 TTACGGTCTCGCCTTAAACCAACGAACGTCG.....CAAG 1307
      ::::::::::::::::::::
315 PheAsnAlaSerSerSerAsnTrpGlnThrThyGlyValIleIleIle 331
      ::::::::::::::::::::
1308 CGCGGGCGCTTCATATCATGATGATGACAGTACCTTACTTGGAAAGTAAAG 1357
      ::::::::::::::::::::
331 rGlnAsnPheSerAlaSerGlyLysSerSerLeuLysPheLysSerGlnG 348
      ::::::::::::::::::::
1358 GCGTGGCAACGACCGCGCTTCCAAATCGCAAAAGGACCGCTGCTGTT 1407
      ::::::::::::::::::::
348 LysThrThrHisAlaAlaPheThrIle.....LysAsnAspLeuIleLeu 362
      ::::::::::::::::::::
1408 CAAGCCAAAGGGGAAAAACCAAGGCTCGTACCGCTGGCGAGCGTAAAGT 1457
      ::::::::::::::::::::
363 AsnAlaThrGly.....GlyAsnI 369
      ::::::::::::::::::::
1458 CATCTTAGATCAGCAGCG.....GACGATCAAGGCAAAAAACAAGCT 1501
      ::::::::::::::::::::
369 eSerLeuAsnGlnValAlaGlyLeaSpSerAsnLeuLysLysSerLeuI 386
      ::::::::::::::::::::
1502 TTAGTGAATCGCTTGTGACGCGGACGAGGAGCGGTGCAACTGATAGCC 1551
      ::::::::::::::::::::
386 leAlaAsnLysAsnIleThrPheGluGlyGlyAsnIleThrLeuAlaAla 402
      ::::::::::::::::::::
1552 GATTAATCAGTTCAACCCGCAAAACTCTATTTCGGCTTTCGGCGGAGAG 1601
      ::::::::::::::::::::
403 .....AspLysLysProIleGluIleLysGlyAsnI 413
      ::::::::::::::::::::
1602 TTGGAATTGACGGGCACTTCCTTCGTTCCACCGCATTCAA...ATA 1648
      ::::::::::::::::::::
413 ethrValLysGlnGlyAlaAsnValThrLeuArgSerAlaAsnTrpGlyA 430
      ::::::::::::::::::::
1649 CCGATGAAGGGGATGATTGTCAACCAACATCAAGCAAAACAATCCACC 1698
      ::::::::::::::::::::
430 snAspLysSerAlaLeuSerIleArgGlyAsnValThrAsnLysGlyAsn 446
      ::::::::::::::::::::
1699 GTTACCATTAACAGCAAT.....AAGATATTACTTACAC 1733
      ::::::::::::::::::::
447 LeuThrValThrGlySerAlaIleAsnIleGluLysAsnLeuThrValG 463
      ::::::::::::::::::::
1734 CGGCAAT..... 1740
      ::::::::::::::::::::
463 uGlySerAlaLysPheLeuAlaAsnProAsnTrpSerPheAsnValSerG 480
      ::::::::::::::::::::
1741 .....AACAAACAATTGATAGCAAAAAAGAAATTCCTTACAAAGCT 1782
      ::::::::::::::::::::
480 YrLeuPheAsnGlnGlnGlyLysSerAsnIleSerIleAlaLysGlyGly 496
      ::::::::::::::::::::
1783 TGGTTGGGGAAGAAATGATCAACCAAAAGCAAGCGGCGCTTAATGTGA 1832
      ::::::::::::::::::::
497 AlaIle...PheLysAspIleGluAsnThr...GlySerLeuAsnIleTh 511
      ::::::::::::::::::::
1833 TTACCAACCGGAAGAAAGCGGATGCGACTTACTGCTTCCGCGGACAA 1882
      ::::::::::::::::::::
511 rThrLysSerAspSerAsnHisThrIleIleLysGlyAsnIleThra 528
      ::::::::::::::::::::
1883 ATTTAAACGGCAATATCAGCAACAAACGCAAACTGTTTTCAGCGGC 1932
      ::::::::::::::::::::
528 snArgLysGlyAspLeuAsnIleThrAsn.....AsnGly 539
      ::::::::::::::::::::
1933 AGACGACACCGCGCCTACAAATGATTTAGAACGGGTGCTCAAAAT 1982
      ::::::::::::::::::::
540 AspAsnThrGlu.....IleGlnIleGlyGlyAsnIleSerGlyL 553
      ::::::::::::::::::::
1983 GGAAGGTATCCCAAGAGAAATGCTGTGGGCAACGATGATGATGACG 2032
      ::::::::::::::::::::
553 sGlu.....GlyAsnLeuThrIleSerSerAspLysValAsnI 566
      ::::::::::::::::::::
2033 GCACATTTAAAGCGGAAACTTCATATTCAAGGCGGA..... 2070
      ::::::::::::::::::::

```

566 IeThr.....GIuargIleThrIleLysAlaGlyValAsnGlyAsp 579  
2071 .....CAAGCGTGGTTCCCGCATGTTGCC..... 2097  
580 AsnSerAspSerAsnGluAlaThrSerAlaAsnLeuThrIleLysThrIly 596  
2098 .....AAAGTGAAGCGATTTGCGATTTAGCAATCAAGCCCAAGCAG 2140  
596 sGIuLeuLysLeuThrAsnAspLeuAsnIleSerGlyPheAsnLysAlaG 613  
2141 TTTTCGGTGTGCGACCGCATCAAGCCACAAATCTGTACAGCTTGGAC 2190  
613 IuIleThrAlaLysAspAsnSerAsnLeuThrIleGlyAspAsnSerAsp 629  
2191 TGG.....ACGGGTGACAGATGTGACGAAAAAACATTAACCGCA 2234  
630 AlaGlyAsnThrAspAlaLysLysValThrPheSerAsnValLysAspSe 646  
2235 TAAAGTATTGCTTCA.....TTGACGACAGACCGACA 2266  
646 rLysIleSerAlaSerAspHisAsnValThrLeuAsnSerLysValGluTr 663  
2267 TCAGAGCAATGTACGCTTCCGATCAGCCTCATTTAAATTCACAGCA 2316  
663 hrSerGlyAspThrAspSerThrGluAspGlyLysAsnAsnThrGly 679  
2317 CTT.....GCCACCTCAGCGCAATCTAGTGACG 2348  
680 LeuThrIleThrAlaLysAsnValThrValAsnAsnIleThrSerHl 696  
2349 CGGAGACAGCGCATATACGTTACGCGCAACGCCACCAAGGCAACC 2398  
696 sLysThrValAsnIleThrAlaSerGluAsnValThrThrLysAlaGly 713  
2399 TCACCTC.....GTGGCAATGCCCAAGCAATTAATCAAGCC 2439  
713 hrThrIleAsnAlaThrThrGlySerValGluValThrAlaLysThrGly 729  
2440 ACATTAAACGCG.....AACACATCGGC 2462  
730 AspIleLysGlyGlyIleGluSerAsnSerGlyAsnValAsnIleThrAl 746  
2463 TTCGAGCAATGCTTCATTAATCAAGCAACACGCGTACAAACGCGCA 2512  
746 aserGlyAsp...ThrLeuAsnValSerAsnIleThrGlyLysValTr 762  
2513 GTCGAGCGCTTCGACACAGCCTAAGCAACGTAAGCATTCGCACTC 2562  
762 hrValAlaAlaAlaSerGlyAlaValThrThrThrLysGlySerThrIle 778  
2563 AAC.....GGCAATGCTCCTAGCGCATTAAGCAGATTCATTT 2603  
779 AsnAlaThrThrGlyAsnAlaAsnIleThrThrIlys..... 790  
2604 TGAAGAACCGCGCTTACCGGAAAAATCAGCGCGGCAAGATACGCA. 2652  
791 .....ThrGlyGluIleAsnGlyGluValLysSerAlas 802  
2653 .....TTACACTTAAAGACAGCGAATGACGCTCGCGTGGGACG 2694  
802 ergLysAsnValAsnIleThrAlaSerGlyAsnThrLeu..... 814  
2695 GAATTAGGCAATTTAACTTGACACAGCCACATTAACACTCAATTCGCG 2744  
815 AsnValSerAsnIleThrGlyLysAsnValThrValThrAlaAsnSerG 831  
2745 CTATTCAGACGATGCGGACGCGCAACCGGACGTCGCGCAATGCGC 2794  
831 Y.....AlaIleThrThrThrGluGlySerThrIleAsnAla. 843  
2795 CGGCGCGCGCTTCGCGCGCTCCCTATTATTCGTTACGCCCAACTTCG 2844  
844 .....ThThr 845

2845 GCAGATCCCGTTTCACACGCTGACG.....GTAAAGCGCAATTGAA 2888  
846 GlyAspAlaAsnIleThrThrGluThrGlyAsnIleAsnGlyLysValG 862  
2889 CGGTACAGGACATTCGCGCTTATGTGCGAACTCTTGGCTACCGCAGC 2938  
862 userSer.....Serg 866  
2939 GCAAATGACGTGCGGAAAGTTCCGAAGCACTTACCTGCGCTGC 2988  
866 LysValThrLeuIleAlaThrGlyGln.....ThLeuAlaVal 879  
2989 AACAAATACCGGCAAGAACCCGTAAGTCTGACGAATTAAGCAGTACGA 3038  
880 GlyAsnIleSerGlyAspThrValThr.....IleThrAlaAspLys 893  
3039 AGGAAAGACACACACCGCTGTCCGAAATCTTAATTCACCTGCGCAA 3088  
893 sGlyLysLeuThrThrGlnThrSerSerLysIleAsnGlyThrLys.Ser 909  
3089 ACGAACAGTGCATGCCGCG.....CATGACGTTATCAGCTATTC 3129  
910 ValThrThrSerSerGlnSerGlnSerLysIleSerGlyThrIleSerGlyAs 926  
3130 CCGAAAGACGCGAGTCCGCTGCATATCCGTCACAAAGAA...CAG 3175  
926 nThrValSerValSerAlaThrGlySerLeuThrThrGlnAlaGlySerL 943  
3176 ACCTTCCGACAACTCGGCAACCGCGGAGAAACAGACGCGCCG... 3219  
943 yIleGluAlaLysThrGlyGluAlaAsnValThrSerAlaThrGlyThr 959  
3220 .....TGACGGCAAAACAGGC 3236  
960 IleGlyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnThrAs 976  
3237 ACACTTCCGCGCAACACACAGCGGAAAAACAGACGCGCAACGCTTG 3286  
976 pAsnLeuThrIleLysAspGlyAlaArgIleLysAlaThrGlyAlaVal 993  
3287 ACGGCTGATTCGCGCGCGCGCATGCCACGCAAAAGCAAGT... 3333  
993 aThrLeuThrAlaThrGlyGlyThrLeuThrThrGluThrSerSerAsp 1009  
3334 .....CTTGCCGAACCGCGCGCA 3353  
1010 IleThrSerSerAsnGlyIleThrThrLeuThrAlaLysAspSerSerL 1026  
3354 GCGAGCGGGGAAATGCC.....GCCATTA 3379  
1026 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1043  
3380 TGCAGCGGAGGAAGAGAAAAACGGGTGACGCGAATTAAGACACCGCC 3429  
1043 euthrThrValAlaGlySerLys...IleGluAlaAlaSerGlyThrIleu 1058  
3430 TTGGCGAAACAGCGCGAAGCGGAACCGCGCGCTACCAACGCTTCC 3479  
1059 ValIleAsnAlaLysAspAlaGlnLeuAspGlyAlaAlaSerGly.... 1073  
3480 CCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCGCGCAACCGC 3529  
1073 ..... 1073  
3530 AACCCACCGCAGCGCAGCTGATCAGCGCTTATGCCAATGACGCTTG 3579  
1074 .....AspHisThrValValAsnAlaThrAsnAlaAsnGlySer 1086  
3580 AGTGAATTTCCGCAACGCTCAACACGCTTTCCGCTACAGACGAAT 3629  
1087 GlySerValIleAlaThrThrSerSerArgValAsnIleThrGlyAspLe 1103



```
219 ..... 219
208 alaSplileuMeIGlnAlaIargGlnGlnsntRP..... 219
859 TTGCAACAGGCAACCCCTATATAGCAAAAAAGCAATGGCTTCAGCTAGT 908
220 ..... As 220
909 TCCTAAAGATTGGTTCTATGATGAATCTTGGTGAAGATACCATTTCAG 988
220 naGATgHIStrP..... GlyATgSerHIStrPA 230
959 TATTCCTGCAACCAATCAAAATGGCAATATCTTTTAAAC..... 999
230 snValThrArgLeuAsnValSerGlnsntSerTyRpheAsnValThrIle 246
1000 ..... GACAATATATATGGCGCAGAGAAAAATCGATGCCAA 1034
247 AspSerSerGlySerAlaSerSerProGlyAlaGlyProLeuAsnAlaG1 263
1035 ACATAAACACTATTCCTCTACCTTATAGATTAATAAAGCAACGACGTTCAAT 1084
263 nSerGlyLeuAsnGlyIleSerPheAsnAsnAspThr..... V 276
1085 TGTTAAATGTTCTTTATCCGAGACAGACAGAACTGTTTATCAATGCT 1134
276 alPheAsnIleAla..... Ala 281
1135 GCAGGTGGGTCAAC... AGTTATCGACC..... AGACTGAA 1169
282 SerSerAlaValAsnPheAsnIleLysProIleValAspLysValTh 298
1170 TAATGGAAGAAATATTTCTTTATGACAAAGAAAGTGAATGTATAC 1219
298 rAsnGlyAsnHisThrLeuPhe..... LysGlyAsnIleSerV 311
1220 TTACCACCAACATCAACCAAGCGCGGCTTTGATTTTGAAGGTAAT 1269
311 alLeu..... GlyGlyLysAlaValAsnPhe..... His 320
1270 TTTACGGCTCGCCTAAACAAACGAACGATG..... CAAG 1307
321 PheAsnAlaSerSerAsnTyRglnThrIleValIleIleGluSe 337
1308 CGCGGGGCTCATATCAGTGTAGTGACGTACCGTACTTGGAAGTAACG 1357
337 rGlnAsnPheSerAlaSerGlyLysSerSerLeuLysPheLysSerLug 354
1358 GCGTGGCAACGACCGCCTGTCCAAATCGGCAAGGACGCTGCTGTT 1407
354 LysThrHisAlaIleAlaPheThrIle..... LysAsnAspLeuIleLeu 368
1408 CAACGCCAAGGGGAAACCAACGAGCTCGGTACGCTGGCGCAGCTAAAGT 1457
369 AsnAlaThrGly..... GlyAsnIle 375
1458 CATCTTAGATCAGCAGGCG..... GACGATCAAGGCAAAAAACACGCT 1501
375 eSerLeuAsnGlnValAlaGlyIleAspSerAsnLeuLysSerLeuI 392
1502 TTAGTGAATCGGCTGTGTCAGCGGACGAGGAGCGGTGCAACTGAATGCC 1551
392 leAlaAsnLysAsnIleThrPheGlnGlyLysAsnIleThrLeuAlaAla 408
1552 GATATATGCTCAACCCGACCAACTATTTGCGCTTTCGCGCGGAGC 1601
409 ..... AspLysLysProIleGlnIleLysGlyAsnIle 419
1602 TTGGATTGAACGCGCATTCGCTTTCACCGCATTCAA... AATA 1648
419 eThrValLysGlnGlyAlaAsnValThrLeuArgSerAlaAsnTyRgLy 436
1649 CCGATGAAGGGCGGATGTTCTCAACCAATCAAGACAAAGATCCACC 1698
::: 1698
436 snAspLysSerAlaLeuSerIleArgGlyAsnValThrAsnLysGlyAsn 452
1699 GTTACCATTTACAGCGCAAT..... AAAGATATTACTACAC 1733
453 LeuThrValThrGlySerAlaIleAsnIleGlnLysAsnLeuThrValG1 469
1734 CGCAAT..... 1740
469 uGlySerAlaLysPheLeuAlaAsnProAsnTyRSerPheAsnValSerG 486
1741 ..... AACACAACTCTGGATACCAAAAAAGAAATTCCTACACGCT 1782
486 lYLeuPheAspAsnGlnGlyLysSerAsnIleSerIleAlaLysGlyGly 502
1783 TGGTTGGGAGAAAGATGCAACCAACGAACGAGCGGCTCATATCTGAA 1832
503 AlAlle... PheLysAspIleGlnsntThr... GlySerLeuAsnIleThr 517
1833 TTACCAACCGGAAGAGCGGATCGCACTTTACGCTTTCGCGCGGACAA 1882
517 rThrLysSerAspSerAsnHisThrIleIleLysGlyAsnIleThrA 534
1883 ATTTAAACGCGCATATACACGCAACAAACGCAACCTTTTTCAGCGGC 1932
534 snArgLysGlyAspLeuAsnIleThrAsn..... AsnGly 545
1933 AGACCGACACCCGACCGCTACAAATCATTTAGAGAACGCGGTGTCAAAAT 1982
546 AspAsnThrGlu..... IleGlnIleGlyGlyAsnIleSerGlnLy 559
1983 GGAAGGTATCCCAAGAGAAATCGTGTGGACACATGTGATGCACC 2032
559 sGlu..... GlyAsnLeuThrIleSerSerAspLysValAsnI 572
2033 GCACATTTAAACGGAACCACTCCATATTTCAGGGCGGA..... 2070
572 lThr..... GluArgIleThrIleLysAlaGlyValAsnGlyAsp 585
2071 ..... CAAGCGGTGTTCCCGCAATGTTGCC..... 2097
586 AsnSerAspSerAsnGlnIleAlaThrSerAlaAsnLeuThrIleLysThrLy 602
2098 ..... AAAGTGAAGCGCATTTGCAATTAAGCAATCACGCCCAAGAG 2140
602 sGlnLeuLysLeuThrAsnAspLeuAsnIleSerGlyPheAsnLysAlaG 619
2141 TTTTGGTGTCCGACCGCATCAAGCAACACATCTGACAGTTCGCGAC 2190
619 lueIleThrAlaLysAspAsnSerAsnLeuThrIleGlyAspAsnSerAsp 635
2191 TGG..... ACGGCTGTGCAAGTTGTACCGAAAAAACCATTCAGCAGA 2234
636 AlaGlyAsnThrAspAlaLysLysValThrPheSerAsnValLysAspSe 652
2235 TAAAGTGAATGCTCA..... TTGACGACAGCCGCA 2266
652 rLysIleSerAlaSerAspHisAsnValThrLeuAsnSerLysValGluT 669
2267 TCAGAGCAATGTACGCTTCCGATACGCTCATTTAAATCTCAGAGA 2316
669 hSerGlyAspThrAspSerThrGlnAspGlyGlyAsnAsnAsnThrGly 685
2317 CTT..... GCCACACTCAACGGCAATCTTATGTGCAGG 2348
686 LeuThrIleThrAlaLysAsnValThrValAsnAsnIleThrSerH1 702
2349 CGAGACACGCACTATACGTTACGCGCAACGCCCAACGCAACGCCAAC 2398
702 sLysThrValAsnIleThrAlaSerGlnsntValThrLysAlaGlyT 719
2399 TCAGCCTC..... GTGGCAATGCCCCAAGCAACATTTATCAAGC 2439
719 hThrIleAsnAlaThrThrGlySerValGlnValThrAlaLysThrGly 735
```

```

2440 ACATTAAAGGC.....AACACATGCGC 2462
      :|||
736 Aspileysgllyllegluserasnsercylasnvallasnilethral 752
      :|||
2463 TTCGACATCTTCATTATTAATCAAGCAACACCGGTACAAAAGGCA 2512
      :|||
752 aserglyasp...Thleuasnvalserasnilethrglylnasnval 768
      :|||
2513 GTCGACGCTTTCGACACGCTTAAGCAACGTAAGCATTCCGCACTC 2582
      :|||
768 hrvallalalalaserglyalvalthrthrthrglyserthrile 784
      :|||
2563 AAC.....GGCAATGCTCTCCCTAGCGGATAGGACAGATTCATT 2603
      :|||
785 Asnalathrthrthrglyasnalaasnilethrthrls..... 796
      :|||
2604 TGAATAACAGCGCTTACCGGAAAAATCAGCGCGGCAAGATACGCA. 2652
      :|||
797 .....Thrglylnleasnnglygluvallysserlals 808
      :|||
2653 .....TACACTTAAAGACAGCAATGAGCGTCCGCTGGGACG 2694
      :|||
808 erglyasnvalasnilethralasercylasnthrleu..... 820
      :|||
2695 GAATTAAGCAATTTAAACCTTGACACGCGCATTAACCTCAATTCGC 2744
      :|||
821 Asnvalserasnilethrglylnasnvalthrvalthralsnsergl 837
      :|||
2745 CTATCGACAGATCGCGGACGCGCAACCGGACGTGGCGAGATCGC 2794
      :|||
837 y.....Alallethrthrthrglyserthrileasnala. 849
      :|||
2795 CGCGCGCGCTGCGCGCTTCCCTATTATCCGTTACGCGCAACTGC 2844
      :|||
850 .....Thrthr 851
      :|||
2845 GCAGATCCGCTTCAACACGCTGACG.....GTAAACGGCAAAATGAA 2888
      :|||
852 Glyaspalasnilethrthrthrglylnleasnnglylvalgl 868
      :|||
2889 CGGTACAGGAACATTCGCTTATGTCGAACTCTCGCTACCGCAGCG 2938
      :|||
868 userSer.....Ser 872
      :|||
2939 GCAAAATGAAGCTGGCGAAAGTTCCGAAGCACTTACACCTGGCGTGC 2988
      :|||
872 lYserValthrleullealthrlyglIn.....ThleuAlaval 885
      :|||
2989 AACAAATACCGGCAACGAAACCGTAAGTCTCGAGCAATGACGTAAGTGA 3038
      :|||
886 glyasnleasercylasprthrvalthr.....llethralasp 899
      :|||
3039 AGGAAAAGACACACACCGCTGTCGAAAAATCTAATTCACCCCTCAAA 3088
      :|||
899 sglylYserthrthrthrglylnleasnnglylYser 915
      :|||
3089 ACAGACACGCTGATGCCGCG.....CATGGCTTATCATCCTTATTC 3129
      :|||
916 ValThrThrserSerInsercylaspilasercylThrilesercylas 932
      :|||
3130 CGCAAAAGCGGCGATTCGCGCTGCATTAATCCGCTCAAGAA.....CAAG 3175
      :|||
932 nThrValSerValSerAlathrthrglyserleuthrthrlnalaglYser 949
      :|||
3176 AGCTTTCGCAAACTCGCAAGCGGAGAAACAGAGCGCGC..... 3219
      :|||
949 yslleaglualalysThrnglygluvalasnvalThrserlathrthr 965
      :|||
3220 .....TTGACGCAAAACAGCGC 3236
      :|||
966 lleglYglYThrilesercylasnThrValasnvalThrAlasnThr 982
      :|||
3237 ACACTTGCCGCCAACACAGCGGAAAAAGCAACAGCGCAACGCTTG 3286
      :|||
982 pasnleuthrThrleYaspolyAlarglleYsalatrhrnglylYalav 999
      :|||
3287 ACAGCTGATTTGGCGCGGCGCAATGCGCAACCAAGCAAGT... 3333
      :|||
999 althrleuthrAlatrhrnglylYThrleuthrThrcluthrserasp 1015
      :|||
3334 .....GTGCCGAACCGCGCGCA 3353
      :|||
1016 lleThrSerSerasnnglylnthrthrthrleuthrAlalYaspSer 1032
      :|||
3354 GCGAGCGCGGGAATGCC.....GGCATA 3379
      :|||
1032 eAlaglySerlleasnAlalasnvalThrleuasnthrthrglyThrL 1049
      :|||
3380 TGCAGCGGAGAGAGAAAAACGCGTCCAGCGGATTAAGACACCGCC 3429
      :|||
1049 euthrthrvalAlaglyserlYs...llegluAlalasercylThrleu 1064
      :|||
3430 TTGGCAAAACAGCGCGACAGCGGAACCGCGCGCTACCAACCGCTTCC 3479
      :|||
1065 ValilleasnAlalYaspAlaglnleuaspGlyAlalasercyl... 1079
      :|||
3480 CGCGCGCGCGCGCGCGGATTTGCCGCAACCGCGCAACCGC 3529
      :|||
1079 ..... 1079
      :|||
3530 AACCCCAACGCGACGCGACCTGATCAGCGTATGCAATAGCGTTTG 3579
      :|||
1080 .....AsphistrValValasnAlatrAsnAlasnnglyser 1092
      :|||
3580 ACTGAATTTCCCGCCACGCTCAACAGCGTTTCCGCTACAGAGCAATT 3629
      :|||
1093 GlyserValillealatrthrthrserSerArYalasnilethrglysp 1109
      :|||
3630 GGAC.....CGCGTGTTCGCGAGACGCGCGCAACGCGC 3664
      :|||
1109 ullerthrleasnnglyleuasnileleaserlYasnnglylYasnthr 1126
      :|||
3665 TTTGGACACGCGCATTCGG...GACACCAACACTACCGTTGC..... 3705
      :|||
1126 allleuYserglYalgluileaspvallystyrlleglp 1142
      :|||
3706 .....CAGATTTCCGCGCTACCGCAACAC 3734
      :|||
1143 AlaserValasnnglyvalilleglualalYasArYalaleuGlYsVally 1159
      :|||
3735 CGACCTGCGCAAAATCGGTATGCAAAAAACCTCGCGCGCGCGCTGC 3784
      :|||
1159 saspileu.SeraspgluualargluthrthrleuAlalYleuGlYalaser 1175
      :|||
3785 GCATCTGTTTTCGCACAAACGCGGAGAACACCTTCGAGCGCATC 3834
      :|||
1176 AlavalArYpHe.....AlaglupRoasnAlalatrThr 1188
      :|||
3835 GGCACCTGCGCGGCTTGCACAGCTGCGCTTTCGGGCAATACGCAT 3884
      :|||
1188 easnThrnglnsnnglylnpethrthrthrgProleuserglInvalThriles 1205
      :|||
3885 CGCAGGT 3892
      :|||
1205 erglucly 1207
      :|||
seq_name: /SIDS1/9cgdate/geneseq/geneseqp-emb1/AA1993.DAT:AA1731
seq_documentation_block:
ID_AA1731 standard; Protein; 1338 AA.
XX_AA1731:
XX_AC
XX_DT 26-APR-1994 (first entry)
XX_XX

```



DE High molecular weight protein 3 (HMW3).  
 XX HMW; high molecular weight protein; virus; vaccine; influenza;  
 KM epitope; immunity; haemophilus influenzae.  
 XX Haemophilus influenzae.  
 OS WO319090-A.  
 XX  
 XX  
 XX PD 30-SEP-1993.  
 XX PF 16-MAR-1993; 93MO-US02166.  
 XX PR 16-MAR-1992; 92GB-0005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 XX PI Barenkamp SJ;  
 XX DR WPI: 1993-320683/40.  
 XX DR N-PSDB; AAO49510.  
 PT High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties  
 XX  
 PS Claim 5: Figure 10; 100pp; English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 SQ Sequence 1338 AA;

alignment\_scores:  
 Quality: 308.50 Length: 1542  
 Ratio: 0.439 Gaps: 75  
 Percent Similarity: 45.525 Percent Identity: 20.169

alignment\_block:  
 US-09-303-518D-653 x AAR41731 ..

Align seg 1/1 to: AAR41731 from: 1 to: 1338

```

1 ATGAAACACACCGACAAACGACACCGAAACACCGCAAGCCCTAA 50
  ||| ||| ||||| ||| :|||
67 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 83
51 AACCGCGCGACCTCCGCTTCGCCCGCTTACTAGCATATGCTGTGCT 100
  :|||:||||:
83 uSerGlyLys..... 86
101 TCGGCATTTCGCCCAAGCCGCGGACACACTTATTCGGCANTCAAC 150
  :|||:||||:||||:||||
87 .....GluGlyGlyGluThrTyrLeuGly..... 94
151 TACCAATACATCCGACCTTTCGCGAAATTAAGCAATTTCGACGTCGG 200
  :|||:||||:||||:||||
95 .....GlyAspGluArgGlyGlu.....G1 101
201 GCGCAAGATATTCAGGTTTACACAAAGGAGGAGTGGTGGCAAT 250
  | :|||:||||:||||
101 YLysAsnGlyIleGlnLeuAlaLysLys.....ThrT 112
251 CGATGACGAAAGCCCGATGATTGATTCTGTGTATCGCGTAAGCGC 300
  :|||:||||:||||:||||
112 hIreGluLysGlySerThrIleAsnValSerGlyLysGluLysGly 128
301 GTGGCGCATTCGGCGGCGCATCAATATATTTGTG..... 333
  ||| :|||:||||| :|||:

```

```

129 ArgAlaIleValTrrpGlyAspIleAlaLeuIleAspLysIleAsnAl 145
334 .....AGCGTGGCACAATAACGGCGGC..... 354
145 aGlnGlyLysAspIleAlaLysThrGlyGlyPheValGluThrSerGlyH 162
355 ..TATAACATGTTGATTGTTGTGCGGAGGAGAACCAATCCC..... 393
162 lSTyrLeuSerIleAspAspAsnAlaIleValLysThrLysGluTrpLeu 178
394 ..GATCAGCACCCGCTTTCCTTACCAATTCGTGAAGAATAATTAATA 440
  ||| :|||:||||:||||
179 LeuAspProGlnAsnValThrIleGlnAlaProSerAlaSerAlaGlyVal 195
441 AGCAGGAGCTAAGCGGCATCCTTATGGCGGATATCATATGCGCGGT 490
  :|||:||||:||||:||||
195 uLeuGlyAlaAspArgAsnSerHisSerAlaGlnValIleLysValThrL 212
491 TGCACAAATTTGTCCACAGATGCAGAACCTTGTGAGTACCATTAATG 540
  ||| ||| ||||| |||||
212 eULysLysAsnAsnThrSerLeuThrThrLeuThrAsnThrThrIleSer 228
541 GATGGGTGGAATACCTGATTTAAATTAATACCTGATTCGTTCGAT 590
  :|||:||||:||||:||||
229 AsnLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysLe 245
591 C..... 591
245 uThrValAsnSerSerIleSerIleGluArgLysSerHisLeuLeuH 262
592 .....GGACGAGCAGACAAATATTCGCGGTGATGAGACGAAACCC 633
  ||| ||| ||| |||
262 lSerGlnGlyGlnGlyGlnGlyValGlnIleAspLysAsp..... 276
634 AATAACCGGAAATTCATATCATATTCGAAGCGCATATTCCTGGCTGT 683
  |||:|||||
277 .....IleThrSerGlu..... 280
684 CGGTGCATATACCTTTCGACAAATAGATCAGTGT..... 720
  ||||| |||||
281 .GlyLysAsnLeuThrIleTyr.....SerGlyLysTrpValAspValH 295
721 ..GGCAGCTCAACTTAGTACGCAAAATTAACATACGCCCATATGCT 768
  :|||:||||:||||:||||
295 lLysAsnIleThrLeuGlySer.....Gly 303
769 TTTTACCAACAGAGGCTCATTTGGCGAC.....AG 800
  ||||| |||||
304 PheLeuAsnIleThrThrLysGluLysIleAlaPheGluAspLysSe 320
801 TGGCTCACCAGATGTTATCTATGATGCCCAAAAGCAAAAGTGTTAATTA 850
  :|||:||||:||||:||||
320 rGlyArgAsnAsnLeuThrIleThrAlaGln..... 330
851 ATGGGATATTCGAACAGCAACCCCTATATAGGAAAGCAATGCTTC 900
  |||:||||:||||:||||
331 ..GlyThrIleThrSerGlyAsn.....SerAsnLysPhe 341
901 CAGCTAGTTCGTAAGATGTTCTATGATGAAATCTTGTGCGAGATAC 950
  :|||:||||:||||:||||
342 Arg.....PheAsnAsnValSerLeuAsnSerIle 351
951 CCATTCAGTATTCCTACGAACACATCAAAATGGAAATATCTTTTAACG 1000
  :|||:||||:||||:||||
351 u.....GlyGlyLysLeuSerPheThrA 359
1001 ACAATATATATGCGCGAGAAAA..... 1023
  |||:||||:||||:||||
359 sPheArgGlnAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 375
1024 .....ATCGATGCCAACA 1037
  |||:|||||
376 PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerIleMetLysAl 392

```

```

1038 TAAACACTATCTCTACCTTATAGATTAACAGACCGTT..... 1080
      ||| ||||| ||| |||||
392 aProlysalserTrpPheIyArGAsPlYsGlyArGhrIyTrPAsnV 409
1081 .....CAATGTTTAATGTTTCTTATCC 1104
      ||| ||||| |||||
409 alThrThleuAsnValThrSerGlySerLysPheAsnLeuSerLLeasp 425
1105 GAGACAGCAGAGAACCTGTTATCATCTCGAGTGGGTCACACAGTGA 1134
      ||| ||||| |||||
426 SerThr.....GlySerGlySerTh 432
1135 TCGACCCAGACTGAATAATGAGAAATATTCCTTATGACAAAGGAA 1204
      ||| ||||| |||||
432 rGlyProSerIleArGAsnIaGlu..LeuAsnGlyIleThrPheAsnL 448
1205 AAGGTGAATTGACTTACCAAGCAACATCAACCAAGCGCGGCGTTTG 1234
      ||| ||||| |||||
448 ySalA.....ThrPheAsnIleAlaGlnGlySer..... 457
1255 TATTTGAGGTAATTTTACGTCGCGCT.....AAAA 1289
      ||| ||||| |||||
458 .....ThrAlaAsnPheSerIleIyAsnIleSerIleMetProPheLysE 472
1290 CAAGCAAACTGGCAA.....GGCGCGGCGTTCATATCATGATGATGCA 1333
      ||| ||||| |||||
472 rAsnAlaAsnTyAlaLeuPheAsnGluAsPlLeSerValSerIyGlyG 489
1334 GTACCGTACTTGAAGTAACGCG..... 1359
489 lYSerValAsnPheLysLeuAsnAlaSerSerAsnIleGlnThrPro 505
1360 .....GTGCCAAACGACCGCTGTCCAAATCGCAAAAGCGACGCT 1400
      ||| ||||| |||||
506 GlyValIleIleIySerGlnAsnPheAsnValSerIyGlySerThrLe 522
1401 GCTGTTCAAGCCCAAGG.....GAAACCC 1426
      ||| ||||| |||||
522 uAsnLeuLysIaGlnGlySerThrGluThrAlaPheSerIleGluAsnA 539
1427 AAGCTCGTCAGCGTGGCGGAGGTAAGCATCTTACATCAG..CAG 1473
      ||| ||||| |||||
539 sPlLeuAsnLeuAsnAlaThrGlyLysAsnIleThrIleArGlnValGlu 555
1474 GCGGACGATCAA..... 1485
      ||| ||||| |||||
556 GlyThrAspSerArGValAsnLysGlyValAlaAlaLysLysAsnIleTh 572
1486 .....GGCAAAAAACAAGCTTTAGTGAAA 1510
      ||| ||||| |||||
572 rPheLysGlyLysAsnIleThrPheGlySerGlnLysAlaThrThrGlu 589
1511 TCGGCTTGTCAGCGGAGGAGGTGCACATGATGCCGAAATACAG 1560
      ||| ||||| |||||
589 le.....LysGlyAsnValThrIleAsnLysAsnThrAsn 600
1561 TTCAACCCCGCAAACTCTATTTGCGCTTTCGCGGAGCGATTTGATTT 1610
      ||| ||||| |||||
601 AlaThr.....LeuArGlyAlaAsnPheAlaG 610
1611 GAACGGGACATTCGCTTTCGTTCCACCGCATCAAAATACCGATGAAGGG 1660
      ||| ||||| |||||
610 uAsnLysSerProLeuAsnIle.....AlaGlyA 620
1661 CGATGATTGTCAACCAACAAT..CAAGACAAGAATCCACCGTTACCAT 1707
      ||| ||||| |||||
620 snValIleAsnAsnGlyAsnLeuThrThrAlaGlySerIleIleAsnIle 636
1708 ACAGGCAATAAGATATTACTACACCGCAAT..... 1740
      ||| ||||| |||||
637 AlaGlyAsnLeuThrValSerLysGlyAlaAsnLeuGlnAlaIleThrAs 653

```

---

```

1741 .....AACACAACTTGATACGAAA 1762
653 nTyTrThPheAsnValAlaGlySerPheAspAsnAsnGlyAlaSerAsn 1762
1763 AAGAAATGTGCTTACACAGGTTGGTTGGCGAGACCGCATCAATCATTTA 1812
      ||| ||||| |||||
670 leSerIleAla..ArgGlyGlyAlaLysPheLysAspIleAsnAsnThr 685
1813 AACGGCGGCTCAATCTGAATTACCAACCGGAAGAGCGATGCACTTT 1862
      ||| ||||| |||||
686 SerSer..LeuAsnIleThrThrAsnSerAspThrThrIyArGhrI 701
1863 ACTGCTTTCGGGGGAAACAATTTAAACGGCATATACGCAACAAG 1912
      ||| ||||| |||||
701 e.....IleLysGlyAsnIleSerAsnLysSerG 711
1913 GCAAACTGTTTTCAGCGGAGACCGACCGACCGCATCAATCATTTA 1962
      ||| ||||| |||||
711 LysAsnLeuAsnIleIleAspLysLysSer..AspAlaGlnIleGlnIle 726
1963 GGAACGGGCTGTCAAAATGAGGTATCCCAAGAGAAATCGTGTG 2012
      ||| ||||| |||||
727 GlyGlyAsnIleSerGlnLysGlu.....GlyAsnLeuThrI 739
2013 GGACACGATTTGATGAC.....CGCATTTTAAACCGGAAA 2050
      ||| ||||| |||||
739 eSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAla... 754
2051 ACTTCATATTCAGGCGGAGCAACGCG..... 2076
      ||| ||||| |||||
755 .....GlyValGlnGlyIyArGSerAspSerSerGlnAlaGluAsnAla 769
2077 .....GTGTTTCCCGCAATGTGCCAAAGTGGAAGCGATTTGCAATT 2120
      ||| ||||| |||||
770 AsnLeuThrIleGlnThrLysGlnLeuLysLeuAlaGlyAspLeuAsnI 786
2121 AAGCAATCACGCCAGCAGATTTGCTGTCGACCGCATCAAAAGCCACA 2170
      ||| ||||| |||||
786 eSerGlyPheAsnLysAlaGlnIleThrAlaLysAsnGlySerAspLeu 803
2171 CAATCTGTACAGTTGGAGCTGAGCGGCTGTGACAAGTTGTCCGAAA 2220
      ||| ||||| |||||
803 hrIleGlyAsnAlaSerGly.....GlyAsnAlaAspAlaLys 815
2221 ACCATTCAGCAGATTAAGTATGATGCTTCATTTAGACAGACCGACATCAG 2270
      ||| ||||| |||||
816 LysValThrPheAspLysValLysAspSerLysIleSerThrAsp...G 831
2271 AGGCATGTCAAGCTTGGCCGATCAGGCTCATTTAATTCACAGAGACTTG 2320
      ||| ||||| |||||
831 yHAsnValIThrLeuAsnSerGlu.....ValL 841
2321 CCACACTCAAGCGCAATCTTATGCAAGCGGAGAC.....ACGACATAT 2364
      ||| ||||| |||||
841 yThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThrGlyLeu 857
2365 ACGGTTACG..CGCAAGCGCCCAAAAGGCAACCTCAGCTGTGGG 2411
      ||| ||||| |||||
858 ThrIleSerAlaLysAspValThrValAsnAsnAsnValThr..... 871
2412 CAATGCCCAAGCAATTTAATCAAGCCACATTTAAACGGCAACACATCGG 2461
      ||| ||||| |||||
872 ...SerHisLysThrIleAsnIleSerAlaAlaIaGlyAsnValThr 887
2462 CTTCGGACATGCTTCATTTAATCTTAAGCAACACCGCGTA..... 2502
      ||| ||||| |||||
887 hrLysGlnGlyThrThrIleAsnAlaThrThrGlySerValGlnValThr 903
2503 .....CAAAACGGGAGTCT 2516
      ||| ||||| |||||
904 AlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrVa 920
2517 GACGCTTTCGACAAC.....GCTAAGGCAAGGTAGCCATTCGCGAC 2560

```

```

920 1ThrlAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaT 937
2561 TCACGGCAGCAATGTCTCCCTAGCCGATAGGACAGTATTCATTTTGAAC 2610
937 hrserGlyThrValAsnIleSerThrLys..... 946
2611 AGCCGCTTACCGGAAAAATACAGCGCGGCAAGATACGCA..... 2652
947 .....ThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAs 960
2653 .TTACACTTAAAGACAGCAATGAGAGCTCCGCTGGCGACAGCAATTAG 2701
960 nValAsnIleThrAlaSerGlyAsnThrLeu.....LysValS 973
2702 GCATTTTAAACCTTGACACGCGCATTACTCAATTCGCTTCGA 2751
973 eAsnIleThrGlyGlnAspValThrValThrAlaAsp..... 985
2752 CACGATCGCGCAGCGGCAACCGGACGAGTGCAGATCGCGCGCG 2801
986 .....AlaGlyAlaLeuThrThrThrAlaGly..... 994
2802 CCGTTCGCGCGCTTCCTATTATCCGTTACCGCGCAACTTCGCGAGAAT 2851
995 .....SerThrIleSerAlaThrThrGlyAsnAla..... 1004
2852 CCGCTTCAACACGCTGACGCTTAACGCGCAATTAAGAGCTCAGGACA 2901
1005 .....AsnIleThrThrLysThrGlyAspIleAsnGlyLys..... 1016
2902 TTCGCTTTATGTCGAACCTTCGCTACCGCAGCGCAATTAAGCT 2951
1016 ..... 1016
2952 GCGCGAAAGTTCGAGAGCACTTAC.....ACCT 2980
1017 .ValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrL 1033
2981 TGCGTGTCAACATACCGGACAGCAACCCGTAAGTCTCGAGCAATGACG 3030
1033 euAlaValGlyAsnIleSerGlyAsnThrValThr.....IleThr 1046
3031 GTAGTGAAGAGAAAGACACACACCGCTGCCGAAATCTTAATTAC 3080
1047 AlaAspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyTh 1063
3081 CCTGCAAAACGACACGTCGATGCCGCGCATGGCTTATCAGCTTATCC 3130
1063 T..... 1063
3131 GCAAAAGACGGAGTTCGCTCATATCCGTCAAAGAACAGACCTT 3180
1064 .....AsnSerValThrThrSerSerGlnSer 1072
3181 TCCGCAAACTCGGCAAGCGGAGAAACAGAGCCGCTTGACGGCAAA 3230
1073 GlyAspIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAla 1089
3231 ACAGGCACTT.....GCCGCAAAACAGAGCGGAGAAAG 3268
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla...LysA 1105
3269 ACAACGGCGCAAGCCTTGACGCGCTGATGGCGCGGCGCAATGCAACC 3318
1105 snGlyAlaAlaThrLeuThrAla.....GluSerGlyLysLeuThrThr 1119
3319 GAAAGCGCAGAAAGTGT.....GC 3338
1120 GlnThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAl 1136
3339 CGAACCGCGCGGCGAGCGGAGAAATGCC..... 3372

```

```

1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrLeuA 1153
3373 .....GCCATATATGACGCGCAGAGAGAGAAAAAGCGGTGACAGCG 3414
1153 snThrThrGlyThrLeuThrThrThrGlyAspSerLys...IleAsnAla 1168
3415 GATAAGACACCGCTTGCGAAACAGCGGCAAGCGGAAACCGCGCGCG 3464
1169 ThrSerGlyThrIleThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185
3465 TACACCGGCTTCCCGCGCGCGCGCGCGCGCGGATTTGCCGCAAC 3514
1185 aAlaSerGly..... 1188
3515 CGCACCCCAACCGCAACCCCAACCGCAGCGCATGATGAGCGTTAT 3564
1189 .....AspArgThrValValAlaAlaThr 1196
3565 GCCAATAGCGGTTTGAATTTTCCGCAACGCTCAACAGCGTTTCGC 3614
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerSerValAs 1213
3615 CGTACAGCAGCAATTGAC.....CGCGTGTTCGCGGAG 3649
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1230
3650 ACCGCGCAGACGCGCTTGACAAACGCGCATCCGG...GACACCAACAC 3696
1230 snGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTr 1246
3697 TACCGTTCGAAGATTTCCGCGCTACCGCCCAACACCGACCTCGCGCA 3746
1247 Ile.....GlnPro.GlyValAlaIleSerValGluGluValIleGluValA 1261
3747 AATCGGTATGCAAAAAAC.....TCG 3769
1261 ysArgValLeuGluValLysAspLeuSerAspGluGluArgGluThr 1277
3770 GCAGCGGCGCGCTCGGACATCTGTTTCGCACACCGGACCGGAAACAC 3819
1278 LeuAlaLysLeuGlyValSerAlaValArgPheValGluProAsnAla 1294
3820 TTCGACGAGCGCATGCGCACTCGGACGCGCTGCCCAACGTCGCGTTT 3869
1294 aIleThrVal.....AsnThrGlnAsnGluPheThrThrLysProSerS 1309
3870 CGGCAATACGGCATCGCAGCTTCGACA.....TCGCGCATCAGCGCGG 3913
1309 eArgIleValThrIleSerGlyGlyLysAlaCysPheSerSerGlyAsnGly 1325
3914 GCGCGGCTTTAGTAGCGGCGCAGCG 3937
1326 AlaArgValCysThrAsnValAla 1333
seq_name: /SIDS1/gcgdata/geneseq/geneseq-embL/AA2000.DAT:AA81609
seq_documentation_block:
ID AA81609 standard; Protein; 1237 AA.
XX
XX AA81609;
XX
XX 24-MAY-2000 (first entry)
XX
XX Streptococcus pneumoniae type 4 protein sequence #109.
XX
XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX antibacterial; antiinflammatory; meningitis; infection; diagnosis;
XX pneumococcal disease.
XX
XX Streptococcus pneumoniae.
XX
XX W0200006737-A2.
XX

```











```
2449 GGCACACATCGCTCGGACATGCTTCAATTAATCAAGACAAACGC 2498
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
865 GYlaAsnPhelaInlaAsnLysSerProLeuAsnIleAlaGlyAsnVa 861
2499 CGTCAAAACGGCAGTGTGACGCTTCCGACACGCTAAGCAACGTAA 2548
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
881 LIleAsnAsnGlyAsnLeuThr.....Thra 890
2549 GCCATTCGCGACATCAAC.....GCAATGTCCTAGCGGATTAAGCA 2592
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
890 laGlySerIleIleAsnIleAlaGlyAsnLeuThrValSerLysGlyAla 906
2593 GTATTCATTTTGAACAGCGCGCTTACGGAAATCAAGCGCGGCA 2642
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
907 AsnLeuGln...AlaIleThrAsnLysThrPheAsnValAlaGlySerPh 922
2643 GGATACGGCATTCATTAAAGACACGGAATGACGCTGCGCGGCA 2692
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
922 eASP..... 923
2693 CGGAATTAGCAATTTAACTTGACAAACGCAACATTACATCAATTC 2742
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
924 .....AsnAsnGlyAlaSerAsnIleSerIleAlaAlaGlyGly 936
2743 GCCTATCGACACGATGCGGACGCGCAACCGGACGTCGCGAGATGC 2792
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
937 AlAlaSpheLysAspIle..... 942
2793 GCGCGCGCGCGCGCTGCGCGCTTCCATTAATCGGTACGCGCAACT 2842
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
943 .....AsnAsnThrSerSerLeuAsnIleThr.....Thra 953
2843 CGGACGATCCCGTTTCACACGCTGACGGTAAACGGCAATTAAGCGT 2892
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
953 snSerAspThrThrTyAlaGlyThrIle...IleLysGlyAsnIleSerAsn 968
2893 CAGGGAACATTCGCTTATGTGGAACTCTGCGTACCGACGCGCA 2942
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
969 .....LysSerGlyAs 972
2943 ATTGAACCTGGCGAA..... 2958
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
972 pLeuAsnIleIleAspLysSerAspAlaGluIleGluIleGlyAla 989
2959 .....ACTTCGAGCGCACTTACACTTGGCT.....GTCAAC 2991
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
989 snIleSerGluLysGluGlyAsnLeuThrIleSerSerAspLysValAsn 1005
2992 AATACCGGCAACGAAACCGTAACTCTGACCAATTGACGGTA..... 3033
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1006 IleThr.....AsnGluIleThrIleLysAlaGlu 1015
3034 .GTGGAA..GGAAGACACACACCGCTGTCGGAATCTTAATTCA 3079
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1015 yValGluGlyAlaGlySerAspSerSerGluAlaGluAsnAlaAsnLeu 1032
3080 CCTGCAAAAGCAACACGTGATGCGCGCGCATGGCTTATCAGCTTATC 3129
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1032 hrlleGln..... 1034
3130 CGCAAAAGCGGAGTTCGCTGCTGATATCCGTCAAAGACAAAGACT 3179
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1035 .....ThrLysGluLeuLysLe 1040
3180 TTTCGACAAACTCGGACGCGGAGGAAACAGAGCGCGCTTACGGCAA 3229
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1040 uAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaI 1057
3230 AACAGGACACACTTGCAGCAACACAGGCGGAAAGACACAGCGGCA 3279
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1057 ysaAsnGlySerAspLeuThrIleGlyAsnAlaSerGlyLysAlaAsp 1073
3280 AGC.....CTTGACGCGCGTGAATTGCGGCC... 3303
1074 AlAlaLysValThrPheAspLysValLysAspSerLysIleSerThrAs 1090
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3304 .GGCGGCATGCCACCGAAAGCAGAA...AGTGTGCCACCGCGCC 3349
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1090 pGlyHisAsnValThrLeuAsnSerGluValLysThrSerAsnGlySer 1107
3350 GGCAGCGCGCGGGGAAATGCC...GGCATATGACAGCGGAGGAGACAG 3396
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1107 eTrAsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAlaLysAsp 1123
3397 .....AAAAACGGGTGACGCGCATTA 3419
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1124 ValThrValAsnAsnValThrSerHisLysThrIleAsnIleSerAl 1140
3420 AGACACCGCGCTTGCAGAAACAGCGGACGGAACCGCGCGCTACCA 3469
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1140 aAlaAlaGlyAsnValThrThrLysGluLysThrIleAsnAlaThrT 1157
3470 CGGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3519
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1157 hreLysSerValGluValThrAlaGluAsnGlyThrIleLysGlyAsnI 1173
3520 CCCCAACCGCAACCCCAACCGGACGCGGACGCTGATCAGCGCTTATGCCAA 3569
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1173 eThrSerGluAsnValThrValThrAlaThrGluAsnLeuValThrThng 1190
3570 TAGCGGTTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTGGCGGTAC 3619
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1190 LuAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer..... 1204
3620 AGACGAAATTGACCGCGCTTGGCGGAGACCGCGCAACCGCGTTGG 3669
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1205 .....ThrLysThrGlyAspIleLysGly 1213
3670 A...CAAGCGCATCGGACGACCAACACACTACCGTGC..... 3706
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1213 yIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThrL 1230
3707 .AACATTTCCGCGCGCTTACCGCCCAACACGACCTGGCCCAATCGGTA 3754
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1230 eLysValSerAsnIleThrGlyGluAspValThrValThrAlaAspAla 1246
3755 TGCAGAAACCTCGGACGCGGCGCGCGCATCTGTTTCGACACAC 3804
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1247 GlAlaLeuThrThrAlaGlySerThrIleSerAla.....ThrTh 1261
3805 CGA..... 3808
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1261 rGlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValG 1278
3809 .CGGAAACACCTTCGACGACGCGCATCGGACACTCGGACGCGTTGCC 3856
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1278 LuSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAla 1294
3857 ACGGTGCGGTTTTCGGCAAT.....ACGCATCGGCGAGGTTTC 3894
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1295 ValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGlyL 1311
3895 GACATCGCATCAGCGCGCGCGGTTTATGATACGCGCATCTTCAGA 3944
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1311 sLeuThrSerThrValGlySerThrIleAsnGlyThrAsnSerValThrT 1328
3945 CGGCATCAGAGCAAA.....TCGCGCGCGCGCTGC 3976
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1328 hSerSerThrValGlySerThrIleAsnGlyThrIleSerGlyAsnThrVal 1344
3977 TGCATTACGCGCATTCAGCAAGAT.....ACGCGCGCAGGT 4011
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1345 AsnValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaLysVa 1361
4012 TTTCGCGGATTCGCGCATCGAACCGGACATCGCGCGCATATTTGCT 4061
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```



1073 GAACGCTCAATG.....TTAATGTTTCTTATCCGAG 1107  
348 erglyHisaspLeuSerlleGlyaspVallleValaspAlaGly 364  
1108 ACAGCAAGAGACCT.....GTTATCATGCTGCAGGTGGGT 1145  
365 TrpLeuaspProaspPaspValSerlleGluThrLeuThrSerGlyAr 381  
1146 CAACGATTATGACCCACAGCTGAATTAATGAAAAATATTTCTTTAT 1194  
381 GAsn.....AsnThrGlyLaspGlnGlyThrThr 392  
1195 .....GACAAAGAAAAGCTGAATGATCTTACACAGC 1227  
392 hrGlyaspGlyThrLysGluSerProLysGlyAsnSerlleSerLysPro 408  
1228 AACATC...AACCAAGGGGGGGGCTTGTATTTTGAGGGTATTTTAC 1274  
409 ThrLeuThrAsnSerThrLeuGluGlnlleLeuArgArgGlySerLysAr 425  
1275 GGTCTCGCTAAAAACAACAAACAGTGCAGACGGCGGCTTCATATCA 1324  
425 lAsnllleThrAlaspAsnArglleLysValAsnSerSerlleAsnLeuS 442  
1325 GTGATGCGACT..ACCGTTACTTGGAAAGTAAACGGCGGTG..GCAAC 1368  
442 eraSnglySerLeuThrLeuHisThrLysArgaspGlyValLyslleAsn 458  
1369 GACGGCGCTGCCAAATCGCGCAAGGCAAGCGTGTGTTCAAGCCAAAG 1418  
459 GlyaspIleThrSerAsnGluAsnGlyAsnLeuThrLleLysAlaGlySe 475  
1419 G.....GAAACCAAGGCTCGTCAAGCGGCGGCAAGTAAATCATCT 1462  
475 rTrrValaspValHisLysAsnllleThrLeuGlyThrGly.....PheL 490  
1463 TAGATCAGCAGCGGCGAGAT.....CAAGCAAAAA 1494  
490 euAsnllleValAlaGlyaspSerValAlaPheGluArgGluGlyAspLys 506  
1495 CAACGCTTAGTGAATCGGCTGTGTCAGGGCGAGGGGACGGGCAACT 1544  
507 AlArGysnAlaThrAspAlaGlnlleThrAlaGlnlleThrIleThrVa 523  
1545 GAATGCGGATATCATGTTCAACCCGCAAACTTATTTGGCTGGCG 1594  
523 lAsnLysaspAspLys.....GlnPheArgPheAsn 534  
1595 GCGGACGCTTGGATTGAAGGGGATTCGCTTGTCTCCACCGCATTCAA 1644  
534 snValSerlleAsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln 550  
1645 AATACCGATGAAGGGGATGATGTGCACACACATCAAGACAAGATC 1694  
551 AsnAsn.....PheThrHisLysPheAspGlyLys.. 560  
1695 CACCGTACATTACAGGCAATAAGATATTACTACAAACCGGCAATACA 1744  
561 .....lleAsnllleSerGlylleValThrIleAsnGlnThr..... 572  
1745 ACAACTTGGATAGCAAAAAAGAAATTCCTACACAGCTGTGTTGGCGAG 1794  
573 .....ThrLysLysaspValLysTr.....TrpAsnAlaSer 583  
1795 AAGATGCAACAAACAAAGCAAGGGGCTCAATCTGAATTAACAAACGGA 1844  
584 LysaspSerLysTrpAsnValSerSerLeuThrLeuAsn..... 596  
1845 AGAAGCGGATGCACTTACTGCTTCCGGCGGAACAAATTTAAAGGCA 1894  
596 ..... 596  
1895 ATATCAGCAAAACAAAGGCAAACTGTTTTCAGCGGCGAGACGACGACG 1944  
597 .....ThrValGlnLysPheThrPhe..... 603  
1945 CACGCTACATATCATTTAGAACGGGTGTCTCAAAATATGGAAGTATCC 1994  
604 .....lleLysPheValAs 608  
1995 ACAAGAGAAATGCTGTGGACAAACGATTGATCGACCGCATTTAAAG 2044  
608 pSerGlySerAsnSerGlnAspLeuArgSerSerArgArgSerPheAlaG 625  
2045 CGGAAATCTTCATATTACAGGGCGGACAAAGCGGTGTTCCGCAATGTT 2094  
625 lValHisPheAsnGlylleGlyGlyLysThr..... 635  
2095 GCCAAAGTGAAGGCGATTTGGCATTTAACATACGCCAACGACTTTT 2144  
636 .....AsnPheAsnllleGlyAlaAsnAlaLysAlaLeuPh 647  
2145 CGGTCTCGACCGCATCAAAAGCCACACATCTGTACAGTTGGACTGGA 2194  
647 elLysLeuLysProAsn.....A 653  
2195 CGGGTCTGCAACGTTGTACGAAAAACCATTAACGACGATTAAGTAT 2244  
653 lAlaThrAspProLysLysGluLeuProIleThrPheAsnAlaAsnllle 669  
2245 GCTTCATTGAGCAAGACCGACATCAGAGCAATGTACGCTTGCCGATCA 2294  
670 ThrAlaThrGlyAsnSerAsp.....SerSerValMetPheAs 682  
2295 CGCTCATTTAAATCTCACA.....GACTTGGCACACATCAACG 2332  
682 pIleHisAlaAsnLeuThrSerArgAlaAlaGlylleAsnMetaspSerI 699  
2333 GCAATCTAGTCAGGCGGAGACGACGACTATAGC..... 2367  
699 leAsnllleThrGlyLysLeuAspPheSerlleThrSerHisAsnArgAsn 715  
2368 .....GTTACGGCGCAACGCCACCCAA 2390  
716 SerAsnAlaPheGluLleLysLysaspLeuThrIleAsnAlaThr...G 731  
2391 CGGCACTCAGCGCTGTGGCAATGCCACGACCAATTTAATCAAGCA 2440  
731 ySerAsnPheSerLeuLysGlnThrLysaspSerPheTrpAsnGluTrpS 748  
2441 CATTTAAAGGCAACACATCGGCTTCGACAAATGCTTCAATTAATCAAGC 2490  
748 erLysHis.....AlaIleAsnSerSerHisAsnLeuThr 759  
2491 AACCAAGCGGTACAAAAGCGAGCTGACGCTTCCGACACAGCTAAGGC 2540  
760 .....lleLeuGlyLysAsnValThrLeuGly.....G 769  
2541 AAACGTAACCATTCGACACTCAACGCAATGTCTCCAGCGGATTAAG 2590  
769 yGluAsnSerSerSerSerlleThrGlyAsnIleAsnllleThrAsnLysA 786  
2591 CAGTATTCATTTTGAAAACAGCGCTTACCGGAAAATACGCGGCGGC 2640  
786 la.....AsnValThrLeuGln 791  
2641 AAGGATACGCA.....TTACACTTAAGACAGCATGAGACCT 2681  
792 AlaAspThrSerAsnSerAsnThrGlyLeuLysLysArgThrLeuThrLe 808  
2682 GCGGTGCGGACGGAATTAGGCAATTTAAACCTTGAC.....AAGCGCA 2725  
808 uGlyAsnIleSerValleGluGlyAsnLeuSerLeuThrGlyAlaAsnAlaA 825  
2726 CCATTAACACTATTCGCTATGACACAGATGCGGCAAGGCGGCAAAAC 2775  
:::||||: ||| ::|

```

825 snlleValGlyAsnLeuSerIleAlaGluAspSer.....ThrPheLys 839
2776 GGCAGTGGCGGAGATGCGCCGCCGCTTCGCGCGCTTCCATTATC 2825
      |||:|||||:|||||
840 GlyAlaIleSerAspAsn.....LeuAs 847
2826 CGTTACGGCGGCAACTTCGCGAGATCCGTTTCAACACGCTGACGGTAA 2875
      |||:|||||:|||||
847 nilethr.....GlyThrPheThrAsnA 855
2876 ACGCAAAATTGAAC.....GCTCAGGAGACATTCGCTTTATGTCG 2916
      |||:|||||:|||||
855 snGlyThrAlaAsnIleAsnIleLysGlyValIleValLysLeuGlnGly 871
2917 GAACCTTCGGCTACCGCGAGCGCAAAATGAAGCTGGGCGAAATTCGGA 2966
      |||:|||||:|||||
872 AspIleIle.....AsnLysGlyGlyLeuAsnIleThrThrAsnAlaSe 886
2967 AGGCACCTTACACCTTGCTGTCACAAATACC..... 2997
      |||:|||||:|||||
886 rGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGluLysGlyA 903
2997 ..... 2997
903 spleuAsnIleLysAsnIleLysAlaAspAlaGluIleGlnIleGlyGly 919
2998 .....GGCAGCAACCGGTAGTCTCGAGCAATGAC 3029
920 AsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysValAs 936
3030 GGTAGTGGAA.....GGAAAGACA 3049
      |||:|||||:|||||
936 nilethrAsnGlnIleThrIleLysAlaGlyValIleGluGlyLysAsn 953
3050 ACACACCGCTCGCAAAATCTTAATTCACCTTCGCAAAACGAAACAGCTC 3099
      |||:|||||:|||||
953 spSerSerGluAlaGlnAsnAlaAsnLeuThrIleGln..... 965
3100 GATCGCGCGCATGCGCTTATCAGCTTATCGCAAAAGCGCGAGTCCG 3149
965 ..... 965
3150 CCGTCATATCCGCTCAAAAGACAAGACTTCGACAAACTCGCAAGG 3199
      |||:|||||:|||||
966 .....ThryLysGluLeuLysLeuAlaGlyLysPleuAsnLies 978
3200 CGGGAAGAACAGAGCGCCGCTTGACGCGAAAGAGCGCAACTTCGCCGCC 3249
      |||:|||||:|||||
978 erGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThr 994
3250 AAACAACAGCGCGGAAAGACAAGCGCGCAAGC..... 3282
995 IleGlyAsnAlaSerGlyGlyAsnAlaAspAlaLysLysValIleThrPheAs 1011
3283 .....CTTGACGCGCTGATTCGCCGC...GGGCGCAATCCACCGGAAA 3322
      |||:|||||:|||||
1011 pLysValLysAspSerLysIleSerThrAspGlyHisAsnValThrLeuA 1028
3323 AGCGCAAGAA...AGTGTTCGCAACCGCGCGGACAGCGCGGCGGAAAT 3369
      |||:|||||:|||||
1028 snSerGluValLysThrSerAsnGlySerAsnAlaGlyLysAsnAspAsn 1044
3370 GCC...GGCATTTATGCAGCGGAGAGAAAG..... 3396
      |||:|||||:|||||
1045 SerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAsnValA 1061
3397 .....AAAAACGGTGCAGCGGATAAAGACACCGCTTGCGGAAC 3439
      |||:|||||:|||||
1061 lThrSerHisLysThrIleAsnIleSerAlaAlaLysValIleThrT 1078
3440 AGCGGGAAGCGGAACCGCGCGCTACACCGCGCTCCCGCGCGCGCGC 3489
      |||:|||||:|||||
1078 hrLysGluGlyThrThrIleAsnAlaThrThrGly.SerValGluValThr 1094

```

```

3490 CGGCGCGCGCGGATTTGCCGCAACCGCGAGCCCGCAACCGCAACCCCAAC 3539
      |||:|||||:|||||
1094 rAlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrV 1111
3540 GCAGCGCGACCTGATCAGCGCTTATCCCAATAGCGTTTGAGTGAATTT 3589
      |||:|||||:|||||
1111 alThrAlaThrGlnAsnLeuValThrThrGluAsnAlaValIleAsnAla 1127
3590 CGGCGACGCTCAACACGCTTTTCGCCGCTACAGAGCAATTTGACCGCGG 3639
      |||:|||||:|||||
1128 ThrSerGlyThrValAsnIleSer..... 1135
3640 TTTGCCGAAGACCGCGCAACGCGCTTTGA...CAAGCGCATCCGGA 3686
      |||:|||||:|||||
1136 ...ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyA 1151
3687 CACCAAAACACTACCTGTTCG.....AAGATTTCCGCGCTTACC 3724
      |||:|||||:|||||
1151 snValAsnIleThrAlaSerGlyAsnThrLeuLysValSerAsnIleThr 1167
3725 GCCAACAACCGACCTGCGCGCAAAATGGTATGCAAGAAAACCTCGGCGAC 3774
      |||:|||||:|||||
1168 GlyGlnAspValThrValThrAlaAspAlaGlyAlaLeuThrThrAl 1184
3775 GGGCGCGTGCAGTCCTGTTTCGCAACCGGA..... 3808
      |||:|||||:|||||
1184 agLysThrIleSerAla.....ThrThrGlyAsnAlaAsnIleThrT 1199
3809 .....CCGGAACACCTTCGACG 3826
1199 hrLysThrGlyAspIleAsnGlyLysValGluSerSerSerGlySerVal 1215
3827 ACAGCATGCGCACTCGGACGAGCTTCGCCAGCGCTTCCTTGGGGCA 3876
      |||:|||||:|||||
1216 ThrLeuValAlaThrGlyAlaThrLeuAlaValGlyAsnIleSerLys 1232
3877 T.....ACGCATCGGCGAGCTTCGACATCGCATCGCGCGG 3914
      |||:|||||:|||||
1232 nThrValThrIleThrAlaAspSerGlyLysLeuThrSerThrValGlyS 1249
1249 erThrIleAsnGlyThrAsnSerValThrThrSerSerGlnSerGlyAsp 1265
3962 .....TCGCGCGCGCGTGTGATTCAGCATTCGACGA 3996
1266 IleGluGlyThrIleSerGlyAsnThrValAsnValAlaThrSerThrG 1282
3997 AGAT.....ACGCGCAGGTTTCGCGGATTCGCGATCGA 4031
      |||:|||||:|||||
1282 yAspLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAlaA 1299
4032 ACCGCAATCGGCGCAACGCGCTATTTCGCAAAAGCGGATACCGAT 4081
      |||:|||||:|||||
1299 lathrLeuThrAlaGlu.....SerGlyLysLeuThrThrGln 1311
4082 ACGAAAACGTCAATTCGCGACCGCGCGGCTTGATTCAAACGCTACCGC 4131
      |||:|||||:|||||
1312 ThrGlySerSerIleThrSerSerAsnGly...GlnThrThrLeuThrAl 1327
4132 GCGG 4135
1327 aLys 1328
seq_name: /SIDSI/gcdata/geneseq/geneseq_emb1/AA1997.DAT.AAM30292
seq_documentation_block:
ID AAM30292 standard; Protein; 1601 AA.
XX AAM30292;
AC
XX
DT 14-APR-1998 (first entry)

```

```

XX Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX
XX Non-typeable Haemophilus; high molecular weight surface protein;
XX HMW4; immunogen; vaccine; otitis media.
XX
OS Haemophilus influenzae strain 5.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 372
XX FT /note- "encoded by TCT"
XX FT Misc-difference 400
XX FT /note- "encoded by AAT"
XX
XX PN W0936914-A1.
XX
XX PD 09-OCT-1997.
XX
XX PF 01-APR-1997; 97MO-US04707.
XX
XX PR 01-APR-1996; 96US-0617697.
XX
XX PA (BARE/) BARENKAMP S J.
XX
XX PI Barenkamp SJ.
XX
XX DR WPI; 1997-503038/46.
XX DR N-PSDB; AAT90993.
XX
XX PT High molecular weight proteins of non-typeable Haemophilus
XX influenzae - useful for vaccine production
XX
XX PS Claim 1; Page 97-102; 183pp; English.
XX
XX
XX This protein comprises the high molecular weight surface protein
XX HMW4 (123 KDa) of non-typeable Haemophilus influenzae strain 5 that
XX has the immunological ability to protect against disease caused by
XX a non-typeable Haemophilus strain and is characterised by at least
XX one surface-exposed B-cell epitope that is recognised by monoclonal
XX antibody Ab6. The HMW4 amino acid sequence was deduced from an
XX isolated hmw4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see
XX AAW30294) and HMW3 (see AAW30291) have also been identified. A
XX conjugate comprising HMW4 linked to an antigen, hapten or
XX polysaccharide, and a synthetic peptide of 6-150 amino acids
XX corresponding to at least protective epitope of HMW4 are also
XX claimed. HMW proteins, conjugates and peptides can be used in
XX vaccines, as immunogens for preparation of antibodies and as
XX antigens for detection of these antibodies.
XX
XX Sequence 1601 AA:
XX
XX
XX alignment_scores:
XX      Quality: 296.50      Length: 1502
XX      Ratio: 0.436      Gaps: 69
XX Percent Similarity: 45.273 Percent Identity: 20.040
XX
XX alignment_block:
XX US-09-303-518D-653 x AAW30292 ..
XX
XX Align seg 1/1 to: AAW30292 from: 1 to: 1601
XX
XX 178 AATTAAGGCAAGTTGCAATCGGGGCAAGATTTAGGTTTACACAA 227
XX ||| ||| : : : : : ||| ||| : : |||
XX 199 AsnHisIleuLeuIleuIleuValGly..LysAspLysVal..... 211
XX
XX 228 AAAAGGAGAGTTGTCGCAAAATGATGACGAAAGCCCGATGATGATT 277
XX : : : : : ||| : : : : : : : : :
XX 212 .....AsnLeuIleuLysValLysValLysVal..... 221
XX
XX 278 TTTTGTGGTATGCGGTAACGCG.....GTGGGCGGATTTGGCGGCGAT 321
XX : : : : : ||| ||| : : : : : ||| ||| : : : : :
XX 222 ..GlyValIleSerValAsnGlyLysIleSerLeuLeuAlaGlyGln 227

```

```

322 CAA.....TATATTGTGAG 335
XX : : : : : ||| : : : : :
XX 238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
XX : : : : : ||| : : : : : ||| : : : : :
XX 336 CGTGGCAGCATACGGCGCTATACACATGTGTGATTTGGTGGGAGGAA 385
XX : : : : : ||| : : : : : ||| : : : : :
XX 254 AlaProGluAsnGluAlaIleAsnLeuGlyAspIleAlaLysGlyG 271
XX : : : : : ||| : : : : : ||| : : : : :
XX 386 GCAATCCGATCAGCACCGCTTTCTTACCAAAATGTGAAAGAAATAAT 435
XX : : : : : ||| : : : : : ||| : : : : :
XX 271 Lysn..... 272
XX
XX 436 TATAAGCAGGACGACGACGACGACGACGACGACGACGACGACGAC 485
XX : : : : : ||| : : : : : ||| : : : : :
XX 272 ..... 272
XX
XX 486 GCGTTTGACAAATTTGTACAGATGACAGAACCTGTGAGATGACCAATT 535
XX : : : : : ||| : : : : : ||| : : : : :
XX 272 ..... 272
XX
XX 536 ATATGATGGTGGAATAATCGCTGATTTAAATTAATACCGTATGCTTT 585
XX : : : : : ||| : : : : : ||| : : : : :
XX 273 .....Ileasn.....Val 275
XX
XX 586 CGAATCGGACGACGACGACGACGACGACGACGACGACGACGACGAC 635
XX : : : : : ||| : : : : : ||| : : : : :
XX 276 ArgAlaIleThrIleAsnLysGlyLysLeuSerAlaLysPheValSe 292
XX : : : : : ||| : : : : : ||| : : : : :
XX 636 TAACCGCGAAAGTTTCATATCATATTCGACGCGATTTCTTG.....C 679
XX : : : : : ||| : : : : : ||| : : : : :
XX 292 LysAspLysSerGlyAsnIleValLeuSerAlaLysGlyGluAlaG 309
XX : : : : : ||| : : : : : ||| : : : : :
XX 680 TCGTGGTGGCAATACCTTTGCACAAATGATGACGTT...GTGGCACA 726
XX : : : : : ||| : : : : : ||| : : : : :
XX 309 LulleGlyValIleSerAlaGlnAsnGlnAlaLysGlyLys 325
XX : : : : : ||| : : : : : ||| : : : : :
XX 727 GTCACACTTGGTACGAAAAAATTAACATAGCCATATGTTTATACC 776
XX : : : : : ||| : : : : : ||| : : : : :
XX 326 LeuMetIleThrIleLysPheValThr.....LeuLys 336
XX : : : : : ||| : : : : : ||| : : : : :
XX 777 AACGAGAGGCTCATTTGGCGACAGTGGTACCAATGTTATCTATGATG 826
XX : : : : : ||| : : : : : ||| : : : : :
XX 336 ThrGlyAlaValIleAspLeuSerLys..... 346
XX : : : : : ||| : : : : : ||| : : : : :
XX 827 CCCAAGACAAAGGTGTAATTAATGGGATTTGCAACAGCAACCC 876
XX : : : : : ||| : : : : : ||| : : : : :
XX 347 .....GluGlyGlyGluThr 351
XX : : : : : ||| : : : : : ||| : : : : :
XX 877 TATATAGGA.....AAAGCAATGCTTCACGCTAGT 908
XX : : : : : ||| : : : : : ||| : : : : :
XX 352 TyrLeuGlyGlyAspLysValGlyGlyLysAsnGlyIleGlnLeuAl 368
XX : : : : : ||| : : : : : ||| : : : : :
XX 909 TCGTAAAGATTGTTTATGATGAATCTTTGCTGGAGATACCCATTGAG 958
XX : : : : : ||| : : : : : ||| : : : : :
XX 368 LysLys.....ThrThrLeuGluLysGlySerThrIleAsnV 381
XX : : : : : ||| : : : : : ||| : : : : :
XX 959 TATTCTAGCAACCAATCAAAATGGGAA..... 987
XX : : : : : ||| : : : : : ||| : : : : :
XX 381 Al.....SerGlyLysGluLysGlyGlyArgAlaIle 391
XX : : : : : ||| : : : : : ||| : : : : :
XX 988 .....TACTTTTAAAGCAATTAATATGCGGAGGAA 1022
XX : : : : : ||| : : : : : ||| : : : : :
XX 392 ValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySe 408
XX : : : : : ||| : : : : : ||| : : : : :
XX 1023 AATGATGCGCAACATTAACACTATTTCTTACCTTATGATTAATAAC 1072
XX : : : : : ||| : : : : : ||| : : : : :
XX 408 LysPheAlaLysThrGlyGlyPhe.....ValGluThr 420
XX : : : : : ||| : : : : : ||| : : : : :
XX 1073 GAACGCTCAATTG.....TTAATGTTCTTATCCGAG 1107
XX : : : : : ||| : : : : : ||| : : : : :
XX 420 erGlyHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGlu 436

```



912 Glycylalanylseraspasn.....LeuAs 919  
 2826 CGTTACGCCGCCAATCTTGGCAGAACATCCGTTTCAACAGCGTGCAGCTAA 2875  
 919 nilethr.....GlyThrPheThrAsnA 927  
 2876 ACGGCAAAATGAA.....GTCAGGGAACATTCGCTTTATGTCG 2916  
 927 snGlyThrAlaAsnIleAsnIleGlyValValAlaLysLeuGlnGly 943  
 2917 GAACCTTTCGCTACCGCAGCGCAATTAAGACTGGCGAAAGTTCGCA 2966  
 944 AspIleIle.....AsnLysGlyLeuAsnIleThrThrAsnAlaSe 958  
 2967 AGGCACTTACACTTGGCTGCACAAATAC..... 2997  
 958 rGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGlnLysGlyA 975  
 2997 ..... 2997  
 975 spleuAsnIleLysAsnIleLysAlaAspAlaGluIleGlnIleGlyGly 991  
 2998 .....GGCAAGCAACCCGTAAGTCTCGACCAATTGAC 3029  
 992 AsnIleSerGlnLysGluIleLysLeuThrIleSerSerAspLysValAs 1008  
 3030 GGTAAGTGA.....GGAAGAACA 3049  
 1008 nIleThrAsnGlnIleThrIleLysAlaGlyValGluGlyGlyAspSera 1025  
 3050 ACACACCCGCTCGCAAAATCTTAATTCACCTTCACAAACGACACGTC 3099  
 1025 spSerSerGluAlaGlnAsnAlaAsnLeuThrIleGln..... 1037  
 3100 GATGCCGGCGCATGGCTTATCAGCTTATCCGCAAGACGGCGAGTTCG 3149  
 1037 ..... 1037  
 3150 CCGCATATATCCGTCACAAAGACAGCTTCCGCAAACTCGGCAAG 3199  
 1038 .....ThrLysGlnLeuLysLeuAlaGlyAspLeuAsnIle 1050  
 3200 CGGCAAGAACAGAGCGCGCTTGAAGGCAAAAGGCACTTCGCGCC 3249  
 1050 erGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThr 1066  
 3250 AAACAAACGGCGGAAAGACACACCGCGAAGC..... 3282  
 1067 IleGlyAsnAlaSerGlyGlyAsnAlaAspAlaLysLysValThrPheAs 1083  
 3283 .....CTGACCGCGCTGATTCGCGCC...GGCGCAATGCCACGAAA 3322  
 1083 pLysValLysAspSerLysIleSerThrAspGlyLysAsnValThrLeuA 1100  
 3323 AGCGAGAA...AGTGTGCGCAACCGCGCGAGCGAGCGGGGGAAT 3369  
 1100 snSerGlnValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsn 1116  
 3370 GCC...GGCATTTATGCAGCGGAGGAAAG..... 3396  
 1117 SerThrGlyLeuThrIleSerAlaLysAspAlaThrValAsnAsnIleVa 1133  
 3397 .....AAAAACGGGTGCGAGCGGCAATTAAGACACCGCTTGGCGAAAC 3439  
 1133 lThrSerIleLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrT 1150  
 3440 AGCGCGAAGCGGAAACCGCGCGCTACACACCGCTTCCCGCGCGCGC 3489  
 1150 hLysGlnGlyThrThrIleAsnAlaThrThrGly...SerValGluValTh 1166  
 3490 CGGCGCGCGCGGATTTCCGCAACCGCAACCGCAACCGCAACCGCAAC 3539  
 1166 rAlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrV 1183

3540 GCAGCGCACCTGATACCGCTTATGCAATAGCGGTTGATGAAATTT 3589  
 1183 alThrAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAla 1199  
 3590 CCGCACCGCTACACAGCGTTTTCGCGCTACAGCAAGATTCGACCGCTG 3639  
 1200 ThrSerGlyThrValAsnIleSer..... 1207  
 3640 TTTCGCGAAGACCGCGCAACCGCGTTTGA...CAAGCGCATCCGCGA 3686  
 1208 ....ThrLysThrGlyAspIleLysGlyIleGluSerThrSerGlyA 1223  
 3687 CACCAACACTACCGTTCC.....AAGATTTCGCGCGCTAAC 3724  
 1223 snValAsnIleThrAlaSerGlyAsnThrLeuLysValSerAsnIleThr 1239  
 3725 GCCAACAACCGACCTGCGCAATTCGCTATGCAAAAACCTCGCGAGC 3774  
 1240 GlyGlnAspValThrValThrAlaAspAlaGlyAlaLeuThrThrAl 1256  
 3775 GGGCGCTGCGCATCTGTTTTCGCAACCGCA..... 3808  
 1256 agLysThrIleSerAla.....ThrThrGlyAsnAlaAsnIleThrT 1271  
 3809 .....CCGGAACACCTTCGACG 3826  
 1271 hLysThrGlyAspIleAsnGlyLysValGluSerSerSerGlySerVal 1287  
 3827 ACAGCATCGCAACCTCGCGACGGCTTCCACGCGCATTCAGGCAAA... 3876  
 1288 ThrLeuValAlaThrGlyAlaThrLeuAlaValGlyAsnIleSerGlyAs 1304  
 3877 T.....ACGCGATCGCGAGGTTGACATCGCATCGCGCGG 3914  
 1304 nThrValThrIleThrAlaAspSerGlyLysLeuThrSerThrValGlyS 1321  
 3915 CGCGGCTTTAGTAGCGCGCAACCTTCACGCGCATTCAGGCAAA... 3961  
 1321 erThrIleAsnGlyThrAsnSerValThrThrSerSerGlnSerGlyAsp 1337  
 3962 .....TCGCGCGCGCGCTGCTGCATTCAGCGCATTCAGGCA 3996  
 1338 IleGluGlyThrIleSerGlyAsnThrValAsnValThrAlaSerThrG 1354  
 3997 AGAT.....ACGCGCAGGTTTCGCGGATTCGCGCA 4031  
 1354 yAspLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAla 1371  
 4032 ACCGCATCGCGCGCAACCGCGCTATTTCGCCAAAAGCGGATTCGCCAT 4081  
 1371 lathrIleuThrAlaGlu.....SerGlyLysLeuThrThrGln 1383  
 4082 ACGAAACGTAATATGCCACCGCGCGCTTCGATTCACCGCTACCGC 4131  
 1384 ThrGlySerSerIleThrSerSerAsnGly...GlnThrThrLeuThrAl 1399  
 4132 GCGG 4135  
 1399 aLys 1400  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1193.DAT:AA11728  
 seq\_documentation\_block:  
 ID AAR41728 standard; protein: 1477 AA.  
 AC AAR41728;  
 AC 26-Apr-1994 (first entry)  
 XX High molecular weight protein 2 (HMW2).  
 DE High molecular weight protein; virus; vaccine; influenza;  
 XX  
 KW

KW epitope; immunity; haemophilus influenzae; gene cluster.

XX Haemophilus influenzae.

OS W09319090-A.

XX 30-SEP-1993.

XX 16-MAR-1993: 93MO-US02166.

XX 16-MAR-1992: 92GB-0005704.

XX (BARE/) BARENKAMP S J.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI: 1993-320683/40.

XX N-PSDB: AA049509.

XX High molecular weight surface proteins - of non-typeable

XX haemophilus which exhibit immunogenic properties

XX Claim 4: Figure 4; 100pp; English.

XX The isolation and purification of the high molecular weight protein

XX enables the identification of the major protective epitopes of the

XX protein by conventional epitope mapping. These epitopes can then be

XX synthesised using standard techniques and incorporated into fully

XX synthetic or recombinant vaccines. This sequence is claimed to be

XX the same as that given in AAR41724 (High molecular weight protein 2)

XX although it does differ slightly.

XX Sequence 1477 AA;

XX alignment\_scores:

XX Quality: 287.00 Length: 1452

XX Ratio: 0.407 Gaps: 69

XX Percent Similarity: 48.623 Percent Identity: 20.110

XX alignment\_block:

XX US-09-303-518D-653 x AAR41728 ..

XX Align seg 1/1 to: AAR41728 from: 1 to: 1477

XX 374 GTGGGGGAGGAGCAATC...CCGATCAGCCGCTTTCTT..... 412

XX 42 ValArgHisLeuAlaLeuLysProLeuSerAlaMetLeuLeuSerLeuL 58

XX 413 ....ACCAAAATTGTGAAGAAATTAATTAAGACAGGAGCTAAGCGCA 458

XX 58 YValThrSerIleProGlnSerValLeuAlaSerGlyAsnLeuThrSert 75

XX 459 TCCTTAATGCGCGGATATCATATGCGCGCTTGCACAAATTGTGCACG 508

XX 75 hrlYsMetLysIlePcysSerPheTyrIleLysThrSerAsnLysThrIle 91

XX 509 ATGGAACACCTGTGAGATGACGAGCTATATGATGGATGGTGAATACGCT 558

XX 92 IleArgAsnSerValAspAlaIleIleAsnTrpLysGlnPheAsn.... 106

XX 559 GATTTAATAATATACCTGATCGTGTTC..... 586

XX 107 .IleAspGlnAsnGlnMetValGlnPheLeuGlnLysAsnAsnAsnSera 123

670 TATTCGTGGCTCGCGTGGCAATACCTTTCACAAATATGATCAGGTGG 719

133 YIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThrAl 169

720 TGGCAGCTCAACTAGGTAGCAAAATTAACATATGATGATGATG 769

169 AsnThrLeuAspIleSerAsnGlnAsnIleLys..... 180

770 TTTTACCAACAGAGAGCTCATTTGGCGACAGCTGCTCACCAGTATTATC 819

181 .....AlaArgAsnPheThr 185

820 TATGATGCCCAAAACCAAAAGTGTATTAATGATGGATATGCAACAGG 869

186 PheGlnGlnThrLysAspLysAlaLeuAlaGlu...IleValAsnHisG 201

870 CAACCCCTATATAGCAAAAGCAATGCTTCACCTGATGATTAAGATT 919

201 YleuIleThrValGlyLysAspGlySerValAsnLeuIle..... 214

920 GGTTCATGATGAAATCTTGTGCTGAGATACCATTCATATTTCTACGA 969

214 ..... 214

970 CCACATCAAAATGGGAATACCTTTTACGACAATATATATGCGCAGG 1019

215 .....GlyLysLysVal.....LysAsnGlnG 222

1020 AAAATCGATGCCCAAAACATAAAGTATCTTCTACCTTATGATTAATAA 1069

222 YValIleSerValAsnGlyLysSerIleLeuAlaGlyLysI 239

1070 CACGAACCGTCAATGTTTATGTTTCTTATCCGACACAGCAAGAA 1119

239 IeThrIleSerAspIleIleAsnProThrIleThrLysSerIleAla 255

1120 CCT.....GTTATCATGCTGCAGGTGG 1142

256 ProGlnAsnGlnAlaValAsnLeuGlnLysPheAlaLysGlyLys 272

1143 GGTCAACAGTTATTCAGACCCAGCTGATATGAGAAATATTTCC.... 1188

272 NleAsnValArgAlaIleThrIleArgAsnGlnGlyLysLeuSerAla 289

1189 .TTTATTCACAAAGGAAA...GGTGAATGATATCTTACCGCAATC 1233

289 spSerValSerLysAspLysSerGlyAsnIleValLeuSerAlaLysG 305

1234 AACCAAGGC...GCGGCGGTTGTATTTGAGGGTAAATTTACGCTTC 1280

306 GlyGlnAlaIleLysLysVal.....IleSe 315

1281 GCCTAATAAACACAAACGTCGCAAGCGCGGCGCTTCAATACATGATG 1330

315 PAlaGlnAsnGlnIleAla...LysLysLysLysLeuMetIleThr...G 330

1331 GCAGTACCGTACTTGAAGATTAAGCGGCTGCAACAGCCGCTGTC 1380

330 LysAspLysValThrLeuLys...ThrGlyAlaValIleAspLeuSerGly 345

1381 AAATTCGCAAAAGCAGCGCTGTC...GTTCAAGCCCAAGGAGAAACCA 1427

346 LysGlnGlnGlyLysLysThrLysLeuGlnGlyAspGlnArgGlnGlnGly 362

1428 AGCTCGCTC.....ACGCTGGCGACGCTGAAGTGA 1459

362 sAsnGlyIleGlnLeuAlaLysLysThrSerLeuGlnLysGlySerThrI 379

1460 TCTTACATCAGCAGCGCGAGATCAAGCAAAACACGCTTATGTA 1509

379 IeAsnValSerGlyLysGlnLysGlyLysValArgAlaIleValTrpGlyAsp 395





```

903 IleserIuSerIaIaThrIhphelysGlyThrIhArgAspThrIeuAsnII 919
2965 .GAAGGACCTTACCTTGGCTGTGCACATACCGCAACGACCAACCCGTAA 3013
919 eThrGlyAsnIhphThr.....AsnAsnGlyThrAlaGluIleA 932
3014 GTCTCGACCAATTGACGGTAGTGGAGAGAAAGCAACACACCGCTGCC 3063
932 snIleThrGInGlyValIValIysIeuGly.....AsnValThrAsnAsp 946
3064 GAAATATTATTTACACCTGCAAAAGCAACACGTCGATGCCGGCGATG 3113
947 GlyAspIeuAsnIleThrIhAlaIaLysArgAsnGInArgSerIleI 963
3114 GCGTTATCAGCTTATCCGCAAAAGACGGCGATTCGCGCTCATATTCGG 3163
963 eGlyGlyAspIleIleAsnIlySylSerIeuAsnIleThrAspSera 980
3164 TCAAAAGACAAGAGCTT..... 3180
980 snAsnAspAlaGluIleGInIleGlyIysAsnIleSerGInIysGly 996
3181 .....TCCGACAAATC..... 3192
997 AsnIleuThrIleSerSerAspIlyIleAsnIleThrIysGInIleThrII 1013
3193 .....GCAAGGCGGAGAGAAACAGAGCGCGCTTGACGCAAAACAGG 3235
1013 eIyIeIysGlyIleAspIlyIleAspSerSerSerAspAlaThrSerSna 1030
3236 CACAACTTGGCCGCAACACAGCGGAGAAAGCAACAGCGCAAAAGCTT 3285
1030 IaAsnIleuThrIleIysThrIysGInIeuIys.....LeuThr 1042
3286 GAGCGCGTGAATGGCGCGGCGCAATGCCAACGAAAGGAGAAAGTGT 3335
1043 GIuAspIeuSerIleSerGlyPheAsn.....LysAlaGluIleTh 1056
3336 TGCCGAACCGCGCCGCGAG.....GAGCGCGGGAAGAAATGCGCGCATTA 3379
1056 rAlaIyAspIyArgAspIeuThrIleGlyAsnSerAsnAspIyAsnS 1073
3380 TGCAGCGCGAGAGAGAAAAACGGGTGCAGCGGATTAAGACACCGCC 3429
1073 eGlyAlaGluAlaIysThrValIhThrPheAsnAsnValIyAspSerIys 1089
3430 TTGGCGAAACAG.....CGCGAAGCGGAAACCGCGCGGCG 3464
1090 IleSerAlaAspGlyHisAsnValThrIeuAsnSerIyValIySyrSe 1106
3465 TACCACGCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGGCGCAAC 3514
1106 rSerSerAsnGlyIyArgIuSerAsnSerAspAsnAspThrGlyIeuT 1123
3515 CGCAGCCCAACCGCAACCCACCGACGCGACCTGATCAGC...CGT 3561
1123 hrIleThrAlaIyAsnValGluValAsnIyAspIleThrSerIeuIys 1139
3562 TATGCCAATAGCGGTTGAGATTTTCGCCACGCTCAACAGCGTTT 3611
1140 ThrValAsnIleThrAlaSerGInIyValIhThr.ThrAlaGlySerT 1156
3612 CGCGGTACAGAGATTTGACCGCGGTTTGGCGGAAGACCGCGCAACG 3661
1156 hrIleAsnAlaThrAsnGlyIyAlaSerIleThrThrIySyrThGlyAsp 1172
3662 CCGTTTGACAAGCGGACCGGACACAAACACTACGCTTGC...AA 3708
1173 IleSerGlyThrIleSerGlyAsnThrValSerValSerAlaThrGlyAs 1189
3709 GATTTCCGCGCTACCGCAACAAACGACCTGCGCAATCGGTATGCA 3758
1189 pIeuThrThrIySerGlySerIyIleGluAlaIySerGlyGluAlaA 1206

```

```

3759 GAAAAACCTCGACGAGCGCGCGTCGCATCTCGTTTTCGCAACCGCA 3808
.. ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1206 snValThrSerAlaThrGlyThrIleGlyGly.....ThrIleSer 1219
3809 CCGAAACACCTTCGACGACGACATCGCA..... 3838
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1220 GlyAsnThrValAsnValThrAlaAsnAlaGlyAspIeuThrValGlyAs 1236
3839 .....ACTCGGACGCGCTTGGCGGCGGCGGCGGTTTTCG 3872
1236 nGlyAlaGluIleAsnAlaThrGInGlyAlaAlaThrIeuThrAlaThr 1253
3873 GCAT.....ACGCAATCGGACGAGTTGCACATCGCATCAGCGGCGG 3916
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1253 IyAsnThrIeuThrThrGluAlaGlySerSerIleThrSerThrIySyl 1269
3917 CCGGTTTATGATGCGGACCTTTCAGACGCGCATCAGAGCAAAATCGC 3966
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1270 GInValAspIeuIeuAla.....GInAsnGlySerIleAlaGlySerII 1284
3967 CGCGGCGTCTGCATTCAGCGCATTCAGGCAAGATACCGCGGAGTTTCG 4016
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1284 eAsnAlaIaAsnValThrIeuAsnThrThrGlyThrIeuThrThrValA 1301
4017 CGGATTCGCGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGCAAA 4066
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1301 IaGlySerAspIleIySAlaThrSerGlyThrIeuValIleAsnIaIyS 1317

seq_name: /SIDS1/9c9data/geneseq/geneseq-emb1/AA1394.DAT:AA63506
seq_documentation_block:
ID AA63506 standard; Protein; 1477 AA.
XX
AC AAR63506;
XX
DT 25-JUN-1995 (first entry)
XX
DE Haemophilus high molecular weight protein HMW2.
XX
KW High molecular weight protein; HMW2; protective vaccine; otitis;
XX sinusitis; bronchitis; Hib; ss.
XX
OS Haemophilus.
XX
PN WO9421290-A.
XX
29-SEP-1994.
XX
15-MAR-1994; 94WO-US02550.
XX
16-MAR-1993; 93US-0038682.
XX
(PARE/) BARENKAMP S J.
XX (SEMW/) ST GEME J W.
XX
PI Barenkamp SJ, St GEME JW;
XX
DR WPI; 1994-316665/39.
XX
Q:PSDB; Q72294.
XX
New immunogenic high mol. wt. proteins of non typeable
XX
Haemophilus - useful in protective vaccines
XX
Claim 3; Page 36; 127pp; English.
XX
XX
The HMW2 protein encoded by this sequence is useful in a vaccine to
XX protect against disease caused by non-typeable Haemophilus which are
XX not controlled by H. influenzae type b (Hib) vaccines. The encoded
XX protein can also be used as a carrier for protective Hib
XX polysaccharide (in a conjugate vaccine against meningitis) or for
XX other antigens, haptens, etc.
XX

```

Sequence 1477 AA;

[illegible]

```
alignment_block:
US-09-303-518D-653 x AAR63506
```

Align seg 1/1 to: AAR63506 from: 1 to: 1477

706	AATGATGAGGTGGTGGCAGTCACTTGTGTAGCGAAATAATTAACA	755
707		
708	.....	
165	AsnGlySerThrIleSerThrLeuAspIleSerAsnGlnAsnIleLys..	180
756	TAGCCCATATGCTTTTATACCAACAGGAGCTCTATTGGCGACAGTGGCT	805
181	.....	
806	CACCAATGTTATTCATGATGCCCAACAGCAAGCTTAATTAATGGG	855
807	:::	
808	.....	
181	IaArgAspNheTrpPheGlnGlnThrLysAspLysAlaLeuGlnGlu..	196
856	GTAATGGCAACAGGCAACCCCTATATAGAAAAGCAATGCCCTCCAGCT	905
857		
197	IleValAsnHisGlyLeuIleThrValGlyLysAspGlySerValAsnLe	213
906	AGTTCGTAAGAATGGTGTCTATGATGAATCTTTGCTGGAGATACCATT	955
213		
213	ttle.....	214
956	CAGTATTCTACGACACATCAATAGGAAATACTTTTATACGACAT	1005
215	.....	
215	.....GlyGlyLysVal.....	218
1006	AATATAGCCGACAGAAAATTCGATGCCCAACATTAACACTATTCTAC	1055
219	.....	
219	..LysAsnGlnGlyAlaIleSerValAsnGlyGlySerIleSerLeuLe	234
1056	TTATATGATTAAACACGGAACCGCTTAATGTTATGTTATGTTTATACG	1105
234	uAlaGlyGlnLysIleThrIleSerAspIleIleAsnProThrIleTrpT	251
1106	AGCAGCAGCAGCACT.....	1128
251	yrSerIleAlaIaIaProGlnAsnGluAlaValAsnLeuGlyAspIlePhe	267
1129	CATGCTGCAGGCTGGGTCAACAGTTATGCACCCACACATCATTAATGAGA	1178
268	AlaLysGlyGlyAsnIleAsnValAlaArgAlaIaIaTrpIleValAsnGlnG	284
1179	AAATATTTCC.....	
284	ylLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleValL	301
1220	TTACCGCAACAATCAACCAAGCG..	1266
301	euserIaLysGlnGlyGlnAlaGlnIleGlyGlyVal.....	313
1267	AATTTACGGCTCGCCTTAACAAACAGCAACGTCGCAAGCGGCGGCGT	1316
314	.....	
314	.....IleSerIaGlnAsnGlnGlnAla..	326
1317	TCATATCATGATGGCAGTACCGTTACTGTGCAAGTAACGCGCTGC	1366
326	uMetIleTrp...GlyAspLysValThrLeuLys...ThrGlyAlaValI	341
1367	ACACACGCGCTGCACAAATACGCAAGCAGCGCTGCTG..	1413
341	IeAspLeuSerGlyLysGlnGlyGlyGlnTrpTrpLeuGlyGlyAspLeu	357

```

1414 AAAGGGGAAAAACCAAGGCTCGGC.....ACGCTGG 1445
358 ArgGlyIuIdIyLysasnGlyIleGlnLysValLysThrSerLeuG1 374
1446 CGAAGCGTAAAGTCATCTTGAATCAGCAGCGGACGCATCAAGCAAAAC 1495
374 uLysGlySerThrIleAsnValSerGlyLysGlnLysGlyArgAlaI 391
1496 AAGCGTTTAAATGAAATCGCGTTGGTAGCGGCAAGGGGACGCTGCAC 1545
391 LeValITrGlyAspRIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGly 407
1546 AATGCGCATATGATGATTCACACCCGCAAAATCTATTTCGGCTTCGGG 1595
408 SerGlyAsp.....IleAlaLysThrG1 415
1596 CGGACGTTTGATTTGACACCGGCATTCGCTTCGTTCCACGCATTCAAA 1645
415 yGlyIrrheValGlnThrSerGlyHisTyIleSerIle..... 427
1646 ATACCGGATGAAGCGGAGATGTGTCAACCCAAATCAACAGCAAGAA... 1692
428 ....AspSerAsnAlaIleVal.....LysThrLysGlnTrp 438
1693 .....TCCACCGTTCACAT..... 1707
439 LeuLeuAspProAspAspArgAlaThrIleGlnAlaGlnAspProLeuAlaG 455
1708 ...ACAGCGATTAAGATATTACTACACACCGGCATTAACAAACATTGG 1753
455 nAsnITrGlyIleAsnAspGluPheProThrGlyThrGlyGlnLaseRa 472
1754 ATAGCAAAAGAAAGAAATGCGTCAACACGGTGGTTGGGCAAGAAAGATGA 1803
472 sPrroLysLysAsn.....SerGluLeuLysThr 481
1804 ACCAAACCAACGCGGCGGCTCAATGTGAATTCACAAACCGGAA..... 1845
482 ThrLeuThrAsnThrThrIleSer...AsnTyIleuLysAsnAlaTrpTh 497
1846 .....GAAGCGATCGCATCTTATCGCTTTC..... 1872
497 rIeuAsnIleThrIleAspThrArgLysLeuThrValAsnSerIleAsnI 514
1873 ..... 1875
514 leGlySerAsnSerHisLeuIleLeuHisSerLysGlyGlnArgGlyGly 1920
1876 GGACAAATATTAAAGCGCATATATCGCAAAACAGCGCAACTG..... 1920
531 GlyValGlnIleAspGlyAspIleThrSerLysGlyLysAsnLeuThrII 547
1921 .TTTTTCACCGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1969
547 eTyIrrSerGlyGlyTrpValAspValHisLysAsnIleThrLeuAspGlnG 564
1970 GGTGTCGCAAAATG.....GAAGT..... 1989
564 yPheLeuAsnIleThrAlaAlaSerValAlaPheGlnGlyGlyAsnAsn 580
1990 .....ATCCCAAGAGAGAAATCGT 2009
581 LysAlaArgAspAlaAlaLysIleValAlaAlaGlnGlyThrValTh 597
2010 GTGGGCAACAGCATTTGGATCGACCCGACACTTTTAAAGCGGAAAATCTCCATA 2059
597 rIleThrGlyGlnGlyLysAsp.....PheArgAlaAsnAsnValSerI 612
2060 TTTCAGCG..... 2067
6068 euaSnGlyThrGlyLysGlyLeuAsnIleIleSerSerValAsnAsnLeu 628
.....GCACACGCGTGGTTTC 2084

```

```

629 ThrHisasnleuserglythrIleasnIlesercylAsnIlethrIleas 645
2085 CCGCAATGTTGCCAAGTGGAGAGCATTTAGCAATCACGCC 2134
645 ngIlethrArgIlysnIlethrSerTyrrpGIInthrSer..... 658
2135 AAGCAGTTTCGGTGGCACCAGCATCAACACCAATCTGTACACGT 2184
659 .....HispsperHis..... 662
2185 TCGGACTG.....ACGGTCTGACAACTGTACCGAAAAACCATAC 2228
663 .....TrrpsnvalSerAlaleuasnleuGluThrGlyAlasnpheth 677
2229 CGAGCATTAAGTATTCCTTCATTGAGCAAG.....ACGACATCA 2269
677 rPheIleIystrIleSerSerasnSerIystrGlyleuThrThrGlnTyra 694
2270 GAGCATGTGTCAGCTTGGCATCAGCTCATTTAAATCTACAGACGT 2319
694 rgsrSer.....AlaGlyAlasnphasnGly... 703
2320 GCCACACTCAACGGCAATCTTAGT.....GCAGCGGAGACAC 2357
704 .....ValasnGlyasnMetSerPheasnleuIystrGlyAlaIystrVa 718
2358 GCACATACAGGTTACGGCAACGCCACCCAAAC..... 2391
718 IasnPhelyleuIystrProasnGlyasnMetasnThrSerIystrProleup 735
2392 .....GCCACCTCAGCTCGTGGCAATGCCCAAGCAACA 2427
735 rolleatyrPheleuAlaasnIlethrAlaIystrGlyIystrValPhe 751
2428 TTTAAT...CAAGCCATTAAACGGCAACATCGGCTTCGACAAATGC 2474
752 PheaspIleIystrAlaasnHissercylArgGlyAlaGluIystrMetse 768
2475 TTCATTAACTAAGCAACAACGCCGATCAAAAGCGAGCTGACGGCTT 2524
768 rGluIleasnIleSerasnGlyAla.....AsnphethIleuA 781
2525 CCGACACGCTAAGCA.....AACGTAAAGC 2550
781 snSerHisvalArgAlaaspAlaIystrIleasnIystrAspIleThr 797
2551 CATTCCGACTCAACGGCAATGTCTCCCTAGCCGATTAAGCATATTCCA 2600
798 IleasnIaIthrAsnSerasnPheSerIleuArgIInthrIystrAspAsp 814
2601 TTTTGA.....AACGCCGCTTACCGCAAAA 2629
814 eTyraSpelyTyrrAlaArgasnAlaIleasnSerThrTyraasnIleSerI 831
2630 TCACCGCGGCGAAGATACGGCATTAACCTTAAAGACAGCAATGAGC 2679
831 IeIeuGlyGlyAsnValIthr...LeuGlyGlyIasnSer..... 843
2680 CTGCCGTGGGACGAGATTAGCAATTTAACTTGACAAAC...GCCAC 2726
844 .....SerSerSerIleThrGlyAsnIleThrIleGluIystrAlaIas 858
2727 CATTAACACTCAATCCGCTTATCAACAGATGGCGGACGCCGCAACCG 2776
858 nvalIthrIeu..... 861
2777 GCAGTGGCGAGATGCCCGCGCGCTTCGCGCTCCCTATTTATCC 2826
862 ..GluAlaasnAlaProasnGlnIasnIleArgAspArgVal... 876
2827 GTTACGCCGCAACTTGGGAGAAATCCCGTTTCAACAGCTGACGGTAA 2876

```

```

877 .....IleIystrleuGlySerIleuValas 885
2877 CGCAATTTGAACGGTACGAGACATTCGGCTTATGCGGACTCTTCG 2926
885 ngIySerIeuSerIeuThrGlyGlyIasnAlaaspIle..... 897
2927 GCTACCGAGCGGCGCAATTTGAAGCTGGCGGAAGTTC..... 2964
898 .....LysGlyasnIeuThrIleSerIeuSerAlaIthrPheIystrGly 911
2965 .....GAAGCACTTACACTTGGCTGTCAA 2990
912 IystrIArgAspPthrIeuasnIleThrGlyasnPheThr..... 924
2991 CATACCGGCAAGAACCGGTATGCTGACGAATTAAGCTAGTGAAG 3040
925 ..AsnasnGlyThrAlaGluIleasnIleThrGlnGlyAlaIystrleuG 941
3041 GAAAAGACACACACCGCTGTCCGAAAATCTTAATTCACCTGCACAAAC 3090
941 Iy.....AsnAlaIthrAsnAspGlyAspIeuasnIleThrThrHisAla 955
3091 GAACACGTGATGCGCGGCGCATGGGCTTATCACTTATCCGAAAGACG 3140
956 IystrArgasnGlnArgSerIleIleGlyIystrIleIleasnIystrGly 972
3141 CGAGTTCCGCTGCATTAATCCGTCAAAGACACAGCTT..... 3180
972 IySerIeuasnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGly 989
3181 .....TCGCAACAATC 3192
989 IyasnIleSerGlnIystrGlyIystrIleuThrIleSerSerAspIystrIle 1005
3193 .....GGCAAGCGGCGAAGACAGA 3212
1006 AsnIleThrIystrGlnIleThrIleIystrGlyIleAspGlyIystrSe 1022
3213 GCGCGCTTGAACGGCAAAACAGGCAACATTCGCCCAACACACAGCGG 3262
1022 rSerSerAspAlaIthrSerAsnAlaIasnIeuThrIleIystrIystrGlu 1039
3263 AAAAGACACACCGCAACCGCTTGACCGCTGATTCGCGCGCGGCAT 3312
1039 eulys.....IeuThrGluAspIeuSerIleSerGlyPheasn 1051
3313 GCGACCGAAAGACGAAGTGTGCGGACCGCGCGGAC.....GC 3356
1052 .....LysAlaGluIleThrAlaIystrAspGlyArgAspIeuThrI 1065
3357 AGCGGCGGAAATTCGCGCATTAATGACGCGGAGAGAGAAAGAAACGG 3406
1065 eglIySerSerAsnAspGlyAsnSerGlyAlaGluIystrValIthrP 1082
3407 TGCAGCGGATTAAGACACCGCGCTTGCGCAACAG..... 3441
1082 heasnIasnValIystrAspSerIystrIleSerAlaAspGlyHisasnValIthr 1098
3442 CGGGAAGCGGAAACCGCGCGGCTTACACCGCTTCCCGCGCGCGCG 3491
1099 IeuasnSerIystrValIystrIystrSerSerasnGlyIystrGluSerAs 1115
3492 CGCGCGCGGGATTTGCGCAACCGACCGACCGCAACCGCAACCGC 3541
1115 nserAspAsnAspThrGlyIeuThrIleThrAlaIystrAsnValGluValA 1132
3542 AGCGGACCTGATCAGC...CGTTATGCCAATAGCGTTTGATGATATT 3588
1132 snIystrAspIleThrSerIeuIystrIystrValAsnIleThrAlaSerGly 1148
3589 TCGCGCACGCTCAACAGCTTTGCGCGCTTACGAGCAATTTGACCGCGT 3638
1149 ValIthrThr.ThrAlaGlySerThrIleasnIaIthrAsnGlyIystrAlas 1165

```

```

3639 GTTTCGCAAGACCGCGCAAGCGCGTTTGCAAGCGGCATCCGGACA 3688
      ::      ::      ::      ::      ::      ::      ::
1165 erlleThrlThrlThrlThrlThrlThrlThrlThrlThrlThrl 1181
      ::      ::      ::      ::      ::      ::      ::
3669 CCAAACTACTACGTTCCG...AAGATTCCGCGCTACCGCCACAAC 3735
      ::      ::      ::      ::      ::      ::      ::
1182 ValSerValSerAlaThrGlyAspLeuThrThrlYsserGlySerIy 1198
      ::      ::      ::      ::      ::      ::      ::
3736 GACCTGCGCCCAATCGCTATGCAAGAAAACCTGGCGACGGCGCGT 3785
      ::      ::      ::      ::      ::      ::      ::
1198 eglulAlaysserGlyAlaAlaSnValThrlSerAlaThrGlyThrl 1215
      ::      ::      ::      ::      ::      ::      ::
3786 CATCTGTTTTCGCAACACGACCGCAACACCTTCGACGACGCGAT 3835
      ::      ::      ::      ::      ::      ::      ::
1215 LyGly.....ThrlSerGlySnThrValAlaSnValThrlAla 1228
      ::      ::      ::      ::      ::      ::      ::
3836 GCA.....ACTCGGCACGG 3849
      ::      ::      ::      ::      ::      ::      ::
1229 AlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThr 1245
      ::      ::      ::      ::      ::      ::      ::
3850 CTTCGCCACGGTTCGCGGCAAT.....ACGGCATCGGCAGATT 3893
      ::      ::      ::      ::      ::      ::      ::
1245 yAlaAlaThrLeuThrAlaThrGlyAsnThrLeuThrThrlGluAl 1262
      ::      ::      ::      ::      ::      ::      ::
3894 CGACATCGGCATCAGCGCGCGCGGTTTATGACGCGCAGCTTTCAG 3943
      ::      ::      ::      ::      ::      ::      ::
1262 erSerlleThrSerThrlYsglyGlyValAspLeuLeuAla..... 1276
      ::      ::      ::      ::      ::      ::      ::
3944 ACGGCATCAGAGGCAAAATCCGCGCGCGTTCGCTACGTCATTCAG 3993
      ::      ::      ::      ::      ::      ::      ::
1277 AsnGlySerlleAlaGlySerlleAsnAlaAlaSnValThrlLeu 1293
      ::      ::      ::      ::      ::      ::      ::
3994 GCAAGATACCGCGCAGGTTTCGCGGATTCGCGATCGACGCGCAT 4043
      ::      ::      ::      ::      ::      ::      ::
1293 rThrlGlyThrlLeuThrThrlValAlaGlySerAspIleGlyAla 1310
      ::      ::      ::      ::      ::      ::      ::
4044 CGCAACCGCGCTATTTCGTCCAA 4066
      ::      ::      ::      ::      ::      ::      ::
1310 LyThrLeuValIleAsnAlaIlys 1317
      ::      ::      ::      ::      ::      ::      ::

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT: AAB01825

seq\_documentation\_block:

ID AAB01825 standard; Protein: 1221 AA.

AC AAB01825;

DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain Joyce mature HMW1A protein, SEQ ID NO:28.

KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;

KW non-typable Haemophilus influenzae; NTHI; non-encapsulated;

KW recombinant production; Escherichia coli; antihacterial; vaccine;

KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;

KW detection; diagnosis.

OS Haemophilus influenzae strain Joyce.

PN WO200020609-A2.

PD 13-APR-2000.

PF 07-OCT-1999; 99WO-CA00938.

PR 07-OCT-1998; 98US-0167568.

PR 08-DEC-1998; 98US-0206942.

PA (CONN-) CONNAGHT LAB LTD.

PI Loosmore SM, Yang Y, Klein MH;

WPI: 2000-303789/26.  
N-PSDB: AAA52176.  
Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans - Claim 8; Fig 18A-R; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the *trp* promoter) operably linked to a modified hmwABC operon from a non-typable (non-encapsulated) H. influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus. CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the CC non-typable strains of Haemophilus via hybridisation reactions. The CC present sequence represents a mature HMWA protein from a non-typable CC strain of H. influenzae. CC

Sequence 1221 AA;

alignment\_scores:

Quality:	283.50	Length:	1217
Ratio:	0.490	Gaps:	62
Percent Similarity:	47.576	Percent Identity:	20.542

alignment\_block:

US-09-303-518D-653 x AAB01825 ..

Align seg 1/1 to: AAB01825 from: 1 to: 1221

```

574 CCTGATCGTGTTCGAATCGACGACGACGACAAATTTGGCGTGTGAT.. 621
      |||||:::  |||:::|||||  ::  ||:::|||||
1  ProAspAsnValSerIleAsnAlaGlyThrSerGlu...ArgAsnAspAl 16
      ::      ::      ::      ::      ::      ::      ::
622 .....GAAGACGAAACC.....ATAACCGCG 643
      |||||  |||  |||  :::|||||
16 aserProThrGluAspPheProThrGlyAlaGlyIlyAsnAspAsnProL 33
      ::      ::      ::      ::      ::      ::      ::
644 AAGTTTCATATCAT.....ATTGCAAGGCGCATATTCCTGG 678
      ::      ::      ::      ::      ::      ::      ::
33 yLysAsnAlaHisAsnLysProThrLeuIleAsnThrThrLeuLarg 49
      ::      ::      ::      ::      ::      ::      ::
679 CTCGTCGCTGCGCAATACCTTTCACAAATAGATGAGTGGCAGCAGT 728
      ::      ::      ::      ::      ::      ::      ::
50 IleLeuSerGlyAsnThrPheVal...AsnIleThrAlaArgLysArgIl 65
      ::      ::      ::      ::      ::      ::      ::
729 CAACTAGTAGCGAA...AAATTAACATAGCCCATAT..... 765
      ::      ::      ::      ::      ::      ::      ::
65 eThrValasSerAspIleAsnIleLysAspSerSerIleLeuLeu 82
      ::      ::      ::      ::      ::      ::      ::
766 .....GGTTTTCACCAACGAGGAGCGCATTT 792
      |||  |||  |||  :::|||||
82 rPserGluAsnAspAsnSerSerGlyValAspIleIlyGlyAsnIleThr 98

```

```

793 GGGCAGAGTGGCTCACCAGTTTATCTAT.....GATGC 827
   ::::::::::::::::::::|||
99 SerThrThrIleGlySerLeuThrIleTyrSerSerGlyTrpIleAsp115
828 CCAAAAGCAAAAGTGGTTTAAATGAGGATG..... 861
   ::::::::::::::::::::|||
115 eHisLysAsnIleThrLeuAsnSerGlyLeuLeuAsnIleThrThrLysG 132
862 .....CAAAAGCAGCAACCCCTATATA..... 882
   ::::::::::::::::::::|||
132 InclAspIleAlaPheGluLysGlyAsnAsnProThrIleThrGlyGln 148
883 .....GGAAAAGCAATGGCTCCAGCTA..... 906
   ::::::::::::::::::::|||
149 GlyThrIleThrAlaGlyAsnGlyLysGlyPheArgPheGluAsnAlaSe 165
907 .....GTTGCTAAAGATTGGT 922
   ::::::::::::::::::::|||
165 rLeuAsnGlyIleGlyThrGlyLeuLeuPheAsnIleLysArgAsp.... 180
923 TCTATGATGAATCTTGGTGGAGATACCATTCAGTATCTCTACGAACCA 972
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
181 .....LeuGlyAsnAsnPheGlnIleIle..... 188
973 CATCAAAATGGAAATCTTTTAAAGCAATATAATAGCGGAGGAAA 1022
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
189 .....AsnPhePheAsnGlyThrLeuAsnIleSerGlyLys 200
1023 AATCGAT.....GCCAAACATAAACACTATCTCTACCTT 1057
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
200 sValAsnIleSerMetValIleProLysLysTrpAspTyrSer..... 214
1058 ATGATTTAAAAACACGACCGTT..... 1080
   ::::::::::::::::::::|||
215 ..LysPheArgGlyArgThrTyrTrpAsnValThrHisLeuAsnValSer 230
1081 .....CAATGTTTAAATGTTCTTTATCCGAGACAGCAAGAACCTGT 1124
   ::::::::::::::::::::|||
231 GluGlySerLysPheAsnLeuThrIleAspSerArgLysAspAsp.... 245
1125 TTATCATGCTGCGAGTGGGTCAACAGCTTATCGACCCAGACTGAATATG 1174
   ::::::::::::::::::::|||
246 .....ThrIleGlyThrLeuAsnThr.....ProTyrAsnLeuAsnG 258
1175 GA.....GAAAATATTCCTTATTTGACAAAGAAAGGAGAA 1212
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
258 lYlIleSerPheAsnLysAspThrIlePheAspValLysGlnAsnGlyAla 274
1213 TTGATA.....CTTACGACACATCAACCAAGCGCG 1244
   ::::::::::::::::::::|||
275 ValThrPheAspIleLysAlaProIleGlyValAsnAsnAsnArgAsnLe 291
1245 GGGCGGTTGATTTTGAAGGTAATTTAGCGTCTCGCTTAAAAACAACG 1244
   ::::::::::::::::::::|||
291 uAsnTyrAlaSerPheAsnGlyAsnIleSerValSerGlyLysLysAsnV 308
1295 AAAGCTGG.....CAAGCGCGCGGGGTT 1317
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
308 aThrPheLysLeuLeuAlaSerSerSerThrAlaGlnThrProGlyVal 324
1318 CATATC.....AGTGATGGCAGTACCGTTACTTG 1346
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
325 PheIleAsnSerLysHisPheAsnAlaSerGlyLysSerSerLeuGluPh 341
1347 GAAAGTAACCGC.....GTGGCAAAAGCA...C 1372
   ::::::::::::::::::::|||
341 eArgThrGluGlySerThrLysValGlyPheLeuIleAsnAsnAsnLeuT 358
1373 GCCTGTCAAAATCGCAAAAGGCAAGCGTGTGTTCAAGCAAGGGGAA 1422
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
358 hrLeuAsnAlaThrGlyLysAsnIleSerLeuLeuGlnValGluGlyLe 374
1423 AAC.....CAAGGCTCGTCAAGCGTGGG 1445

```

```

375 AspGlyMetIleGlyLysGlyValValAlaLysLysAsnIleThrPheAl 391
   ::::::::::::::::::::|||
1446 CGACGGTAAGTCATCTTAGATCAGCAGCGGAGACATCAAGCAAAAAAC 1495
   ::::::::::::::::::::|||
391 nGlyLysAsnIleThrPhe.....GlySerLysL 401
1496 AAGCCTTAGTAAGTAATCGCTTGTGACGCGCAGGCGGCGGTCAACTG 1545
   ::::::::::::::::::::|||
401 yAlaIleThrGluIle.....GluGlyAsnAlaThrIleAsnAsn 414
1546 AATGCCGATATCACTTCACCCGACCAACTCTATTTCGCTTCGCGG 1595
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
415 AsnAlaAsn..... 417
1596 CGGACGTTTGATTTGAACGGGCAATTCGCTTTCACCGCATTCATA 1645
   ::::::::::::::::::::|||
418 .....ValThrIleGlySerAspPheAsnAsnIleGlnLysProL 432
1646 ATACCGATGAAGGGCGATGATGTCTCAACCAACATCA.....GACAAA 1689
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
432 euThrIleLysLysAspValIleIleAsnSerGlyAsnLeuThrAlaGly 448
1690 GAATCCACCGTTTACATTTACAGCAATATAAGATTTACTTACACCGGCA 1739
   ::::::::::::::::::::|||
449 GlyAsnValIleAsnIleAsnGlyAsn.....LeuThrValAsnAs 462
1740 TTAACAACACTGTGATTCGCAAAAGAAATGCTTACAC.....GGTT 1783
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
462 nGlyAlaAsnLeuLysAlaIleThrAsnThrPheAsnValGlyGlyL 479
1784 GGTTCGCGCAG..... 1794
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
479 euPheAspAsnLysGlyAsnSerAsnIleSerIleAlaArgGlyLysAla 495
1795 .....AAGATGCAACCAAAAGCAAGCGCGCTTCATTCGATTTACCA 1838
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
496 LysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrThrAs 511
1839 ACCGGAAGAAAGCGATTCGCTTACTCTGCTTCGCGGCAAAATTTAA 1888
   ::::::::::::::::::::|||
511 nSerAspThrThrTyrArgThrIleIleGluGlyAsnIleThrAsnLys 528
1889 ACGGCAATATCACGCCAAACAAAGCAACTGTTTTCAGCGGACAGCCG 1938
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
528 laGlyAspLeuAsnIleIleAspAsnLys..... 537
1939 ACACGCGACGCTACATCATTTAGAACGGGCTGCAAAATGCAAGC 1988
   ::::::::::::::::::::|||
538 ..GlyAsnAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGluG 553
1989 TATCCCAAGAGAAATCGTGTGGACAAACGATTGATCGACCGACAT 2038
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
553 Y.....AsnLeuThrI 557
2039 TTAACCGGAAATCTCATATTACAGGCGGACAAAGCGTTCGCCG 2088
   ::::::::::::::::::::|||
557 leSerSerAspLysIleAsnIleThr...AsnGlnIleThrIleLysLys 572
2089 AATGTTGCCAAAGTGAAGCGGATTTGATTTAAGCAATCACGCCAAGC 2138
   ::::::::::::::::::::|||
573 GlyValAsnLysGluAspSerAspSerSerThrAlaAsnAlaAsnLe 589
2139 AGTTTTCGCTGTCGACCCGATCAAAAGCACAATCTGTACAGCTTCG 2188
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
589 u.....ThrIleLysThrLys.... 594
2189 ACTGACAGGCTGTGACAGTTGTACGAAAAAACCTTCACGACGATAA 2238
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
595 .....GluLeuGlnLeuThrGlyLysPhe 602
2239 GTGATTGCTTCATTGAGCAAGACGACATCAGA..... 2271
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

603 AsnIleSerGlyPheAspLysAlaGluIleThrAlaLysGluAlaAs 619  
2272 .....GGCAAT..... 2277  
619 PLeuIleIleGlyAsnSerAspAsnAsnAlaAlaLysLysV 636  
2278 .....GTCAACCTTGGCGATACGCT 2298  
636 alThrPheAsnGluValLysAspSerLysIleSerAlaGlySerHisAsn 652  
2299 CATTAATATCTCACAGACTTGGCACCTCAACGGCATCTTAGTGA.. 2346  
653 ValThrLeuAsnSerLysValGluThrSerAsnGlyAsnAsnAlaGlu 669  
2347 .....GGCGAGACACGGCATTACGGT.. ACCGGACACGCCA 2383  
669 uSerAsnAsnGlyAspSerThrSerThrIleAsnAlaLysAsnVal 686  
2384 CCAAAACGGCAACCTC..... 2400  
686 hrValAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAlaSer 702  
2401 .....AGCCTCGTGGCGAA 2414  
703 GluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrIleGlySer 719  
2415 TGCCCAAGCAACATTATATCAAGCCACATTAAACGGC..... 2451  
719 rValGluValThrAlaLysThrGlyAspIleLysGlyGlyIleGluSerA 736  
2452 .....ACACATCGGCTTCGGACATGCTTCATTATTAATCTA 2487  
736 snSerGlyAsnValAsnIleThrAlaSerGlyAsp...ThrLeuAsnVal 751  
2488 AGCAACAACGCCGTACAAAGAGGAGCTGACGCTTCCGCAACGCTAA 2537  
752 SerAsnIleThrGlyGluAsnValThrValAlaAlaSerGlyAlaVala 768  
2538 GGCAACGTAAGCATTCCTCCGACATCAAC.....GGCATGTCTCC 2578  
768 lThrThrThrLysGlySerThrIleAsnAlaThrThrGlyAsnAlaAsnI 785  
2579 TAGCCGATAGGAGCATTCCTATTGAAACACGCCCTTACCGGAAA 2628  
785 lThrThrLys.....ThrGlyGlu 791  
2629 ATCAGCGCGGCAAGATACGCA.....TTACACTTAAAGACAG 2669  
792 IleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThrAlaSe 808  
2670 CGAATGGAGCGTGGCGGCGGCAATGAGCAATTAACTTAACTTGACA 2719  
808 rGlyAsnThrLeu.....AsnAsnValSerAsnIleThrGlyGluA 822  
2720 ACGGCACCATTAACATCAATTCGCTATGACACAGATGCGGCGCG 2769  
822 snValThrValThrAlaAsnSerGly.....AlaIleThr 834  
2770 CAACCGGCACTGCGGAGATGCGCGCGCGCTTCGCGCTTCCT 2819  
835 ThrGluGlySerThrIleAsnAla..... 842  
2820 ATTATCGCTTACGGCCCACTGGGCAAGATCCCGTTTCAACAGCGTGA 2869  
843 .....ThrThrGlyAspAlaAsnIleThrThrGluInt 853  
2870 CG.....GTAACGGCAATTTGAACGCTCAGGACATTCGCTTATG 2913  
853 hrGlyAsnIleAsnGlyLysValGluSer..... 863  
2914 TCGCACTCTTCGGCTACCGGAGCGCAATTTGAAGCTGGCGGAAAGTTC 2963  
864 .....SerGlySerValThrLeuIleAlaThrGlu 873

2964 CGAAGGCACTTACACCTTGCTGTCAACAATACCGCAACGACCGTAA 3013  
873 yGlu.....ThrLeuAlaValGlyAsnIleSerGlyAspThrValT 887  
3014 GTCTCGAGCAATTTGACGCTAGTGAGAGAAAGACACACACCGCTGCC 3063  
887 hr.....IleThrAlaAspLysGlyLysLeuThrThrGluIntThrSer 900  
3064 GAAATCTTAATTTACACCTCGCAAAACGACACGCTGACCGGGG... 3109  
901 SerLysIleAsnGlyThrLys..SerValThrThrSerSerGluSerGlyA 917  
3110 .....CATGCGCTTATCAGCTTATCGGCAAAACGCGCGGCTTCGCTGC 3154  
917 spIleSerGlyThrIleSerGlyAsnThrValSerValSerAlaThrGly 933  
3155 ATATATCGGCTCAAGAA...CAAGAGCTTTCGCAACAACTCGCAAGCG 3200  
934 SerLeuThrThrGluAlaGlySerLysIleGluAlaLysThrGlyGluAl 950  
3201 GGGAGAAACAGAGCGCGCC..... 3219  
950 AsnValThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnT 967  
3220 .....TTGACGGCAAAACAGAGCACAACTTGCAGCAACACAGCGG 3261  
967 hrValAsnValThrAlaAsnThrAspAsnLeuThrIleLysAspGlyAla 983  
3262 GAAAGAAACAGACGCGCAAGCCTTGACGCGCTGATTCGCGCGCGCGCA 3311  
984 ArgIleLysAlaThrGlyGlyAlaValThrLeuThrAlaThrGlyGlyTh 1000  
3312 TGCCACCGAAAGGCAAGAGT..... 3333  
1000 rLeuThrThrGluThrThrSerSerAspIleThrSerSerAsnGlyGluIntT 1017  
3334 .....GTTGCCGCAACCGCGCGGACAGCGCGGCAAAATGCC..... 3372  
1017 hrLeuThrAlaLysAspSerSerIleAlaGlySerIleAsnAlaAlaAsn 1033  
3373 .....GGCATTATGACGCGGAGGAGAAAGACAAAACG 3404  
1034 ValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerLys.. 1049  
3405 GGTGACGGCGGATTAAGACACCGCTTGGCGAAACGCGCGGAGGGA 3453  
1050 ..IleGluAlaAlaSerGlyThrLeuValIleAsnAlaLysAspAlaGlu 1065  
seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA01824  
seq\_documentation\_block:  
ID AAB01824 standard; Protein: 1227 AA.  
XX AAB01824;  
AC  
XX  
DT 11-SEP-2000 (first entry).  
DE  
XX Haemophilus influenzae strain Joyce HMW1A protein, SEQ ID NO:26.  
XX  
XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;  
KW recombinant production; Escherichia coli; antibacterial; vaccine;  
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
KW detection; diagnosis.  
OS Haemophilus influenzae strain Joyce.  
XX  
XX W0200020609-A2.  
XX  
XX 13-APR-2000.  
PD  
XX  
PF 07-OCT-1999; 99WO-CA00938.





```

347 eArGTtRtGlUGlYSerThrLysValGlYPheLeuIleAsnAsnAspLeuT 364
      :::::|||||
1373 GCCTGCCAAATCGCAAGCAGCGCTGCTTCACGAAACCCAAAGGCGAA 1422
      |||||
364 hrlEusnAlArhrlYglYAsnIleSerLeuLengIValGlYlIle 380
1423 AAC.....CAAGCTCGTCAGCGTGG 1445
      :::::
381 AspGlyMetIleGlYlYsolYValAlAlAlYsLysAsnIleThrPheAl 397
1446 CGACGGTAAGTCATTCATTCAGTCACAGCGGAGCATGACAGGCAAAAAAC 1495
      :|||:::
397 aGlYglYAsnIleThrPhe.....GlySerLysL 407
1496 AAGCCTTAGTGAATCGCTTGTGACGGGAGGGGAGCGTGCACATG 1545
      ::|||:::
407 yAlAlIleThrluIle.....GlUGlYsAlAlArhrlIleAsn 420
1546 AATGCCGATATCAGTTCACCCGACAAACTATTTGCGCTTTCGCGG 1595
      |||||
421 AsnAlAsn..... 423
1596 CGGAGCTTGGATTGACGGGCAATCGCTTTCCTCCACCGCATTCAAA 1645
      ::|||
424 .....ValThrLeuIleGlYSerAspPheAspAsnIleSglInYsProl 438
1646 ATACGATGAAGGGCGCATGATGTCACACACAAATCAA.....GACAA 1689
      |||
438 euThrlIleLysAspValIleIleAsnSerGlyAsnLeuThrAlAGly 454
1690 GAATCCACCGCTTACCATTTACAGGCAATAAAGATTACTACACCGGCA 1739
      :::::
455 GLYsNValIleAsnIleAsnGlYAsn.....LeuThrValAsnAs 468
1740 TAACAACACTTGGATAGCAAAAAAGAAATTCCTACAC.....GGT 1783
      |||
468 nGlYAlAsnLeuYsAlAlIleThrAsnPheThrPheAsnValGlYlY 485
1784 GGTTCGCGAG..... 1794
485 eupheAspAsnLysGLYsNserAsnIleSerIleAlAlArGlYGlYAla 501
1795 .....AAAGTCACACCAAAAGACAGCGGCGCTCAATTCGAATTCAC 1838
      |||||
502 LysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrTrAs 517
1839 ACCGAGAGAAGCGGATGCGCACTTACTGCTTTCGCGGCAACAATTTAA 1888
      ::::
517 nSerAspThrThrYrArThrIleIleGlYglYAsnIleThrAsnLysA 534
1889 ACGGCAATATCAGCAAAACAGGCAAACTGTTTTCAGCGGAGACCG 1938
      |||||
534 laGlYAspLeuAsnIleIleAspAsnLys..... 543
1939 ACACCGGACCCCTTCAATTCATTAGAGAGCGGCTGCATAAATGAAG 1988
      ::|||
544 ...GLYsNAlAluIleGlnIleGlYglYAsnIleSerGlnLysGlUGl 559
1989 TATCCCAAGAGAAATCGTGTGGACACAGATTGGATCGACCGGACAT 2038
      |
559 Y.....AsnLeuThr 563
2039 TTTAAGGGGAAACTTCATATTCAGCGCGGACAGCGGTGTTCCCG 2088
      ::|||
563 lSerSerAspLysIleAsnIleThr...AsnGlnIleThrlIleLysLys 578
2089 AATGTGCCAAAGGAGGAGCGATTTAGCAATTCAGCCCAAGC 2138
      ::|||
579 GlYAlAsnLysGLYsNspSerAspSerThrAlAlAsnAlAlAsn 595
2139 AGTTTTCGTGTGCGACCGCATCAAGCAACACATCTGTACACGTTCCG 2188
      |||||

```

```

595 u.....ThrIleLysThLys.... 600
2189 ACTGAGCGGTGTGACAAAGTTGTACCGAAAAACCATTCAGCATAA 2238
      |||
601 .....GluLengIleuThrGlYAspLeu 608
2239 GTGATGCTTCATTCAGCAAGCCGACATCA..... 2271
      |||||
609 AsnIleSerGlyPheAspLysAlAGlIleThrAlAlYsGlUGlYAlAs 625
2272 .....GGCAAT..... 2277
625 PleuIleIleGlYAsnSerAspAsnAsnAlAlAsnAlLysLysY 642
2278 .....GTACGCTTGGCATACCGCT 2298
      ::|||
642 AlThrPheAsnGlnValLysAspSerLysIleSerAlAGlySerHisAsn 658
2299 CATTTAATCTCACAGACTTGCACACTGACAGGCAATTTAGTGA.. 2346
659 ValThrLeuAsnSerLysValGlUGlThrSerAsnGlYAsnAspAlAl 675
2347 .....GGCGAGACACGCACATAAGGTT...ACGCGACAGCCA 2383
675 uSerAsnAsnGlYAspSerThrSerLeuThrlIleAsnAlAlYsAsnValT 692
2384 CCCAAAACGGCAACCTC..... 2400
      ||
692 hValAlAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAlAs 708
2401 .....AGCTGCTGGGCA 2414
709 GlAsnValThrThrlYsAlAGlYThrThrlIleAsnAlArhrlIleGly 725
2415 TGCCCAAGCAACATTTATTCAMCCACATTTAAACGCG..... 2451
725 rValGlUGlValThrAlAlYsThrGlyAspIleLysGlYglYlIleGluSerA 742
2452 .....AACACATCGCTTGGACACANCTTCATTAAATCTA 2487
      |||
742 snSerGlyAsnValAsnIleThrAlAlSerGlYAsp...ThrlEusnVal 757
2488 AGCAACAAGCCGCTACAAAACGAGCTGAGCGCTTTCGACAAAGCTTA 2537
      |||||
758 SerAsnIleThrGlyGlnAsnValThrValAlAlAlAlSerGlyAlAla 774
2538 GGCAAACGTAAAGCATTCGCACTCAAC.....GGCAATGTCCTCC 2578
      ::|||
774 lThrThrThrlYsGLYSerThrlIleAsnAlArhrlThrGlYAsnAlAlAsnI 791
2579 TAGCCGATAGCAGATTCATTTTGAACAAGCCGCTTTACGGGAANA 2628
      ::|||
791 lArhThrLys.....ThrlGlyL 797
2629 ATCAGCGCGCAGAGATACGCA.....TTACATTTAAAGACAG 2669
      |||||
798 lIleAsnGlUGlValLysSerLAsSerGlyAsnValAsnIleThrAlAlSe 814
2670 CCAATGACGCTGCGCGGACGGAATTCAGCAATTTAAACCTTGCA 2719
      |
814 rGlYAsnThrlEus.....AsnAsnValSerAsnIleThrGlYlAl 828
2720 ACGCACCATTCACATTCGCGCTATTCGACACAGATTCGCGGAGCGCG 2769
      |||||
828 snValThrValThrAlAlAsnSerGly.....AlAlIleThrThr 840
2770 CAACCGGAGTGCAGAGATGCGCGCGCGCGCTTCGCGCGCTTCCT 2819
      |||||
841 ThrlGlyLysThrlIleAsnAla..... 848
2820 ATTATCGTTACGCGCGCACTTGGCGAGAAATCCCGTTTCACACGCTGA 2869
      |||||
849 .....ThrlThrGlYAspAlAlAsnIleThrThrlGlnT 859

```





697 .....AlaGlyValAsn pheasn gly .....Val 704  
2329 AACGGCAATTTAGT.....GCAGCGGAGACAGCAGCTATAC 2366  
|||||.....  
705 AsnGlyAsnMetSerPheAsnLeuGlyValAlaValAsnPheGly 721  
2367 GGTTACCGCGAACGCCGCCCAAAAC..... 2391  
||| .....  
721 sLeuGlyPProAsnGlnAsnMetAsnThrSerLysProLeuProIleArgP 738  
2392 .....GGCAACCTCAGCCTCGTGGCAATGCCCAAGCAACATTTAAT... 2433  
|||||.....  
738 heLeuAlaAsnIleThrAlaThrGlyGlySerValPhePheAspIle 754  
2434 CAAGCCACATTAAGCGCAACACATCGGCTTCGACATGCTCATTTAA 2483  
|||||.....  
755 TyrAlaAsnHisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAs 771  
2484 TCTAAGCAACAGCGCGTACAAAAGCGGAGCTTGACGCTTCCGACAAAGC 2533  
|||||.....  
771 nIleSerAsnGlyAla.....AsnPheThrLeuAsnSerHisV 784  
2534 CTAAGCA.....AACGTAAGCCATTCCGCA 2559  
|||||.....  
784 alArgGlyAspAspAlaPheLysIleAsnLysAspLeuThrIleAsnAla 800  
2560 CTCACAGCGCAATGTCCTCCTAAGCCGATAGCAGTATTCATTTGAA.. 2607  
|||||.....  
801 ThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheGlyArgI 817  
2608 .....AACAGCGCTTATCCGGAATAATCAGCGCG 2638  
|||||.....  
817 yTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleLeuGly 834  
2639 GCAGAGATAGCGCATTTACCTTAAGACAGCGAATGAGCGCTGCCGTCG. 2688  
|||||.....  
834 LysAsnValThr...LeuGlyGlyLysAsnSer.....Ser 844  
2689 GGCAAGCAATTAAGCAATTAACCTTGACAAAC...GCCACCATTTACCT 2735  
|||||.....  
845 SerSerIleThrGlyAsnIleThrIleGluLysAlaIleAsnValThrLe 861  
2736 CAATTCGCGCTATCGACACGATGCGGCGCAAAACGCGGAGTGGC 2785  
| .....  
861 u.....GluAla 864  
2786 CAGATGCGCGCGCGCGCTTCGCGCGCTTCCTATATCGTTACGCGCG 2835  
|||||.....  
864 snAsnAlaProAsnGlnGlnAsnIleArgAspArgVal..... 876  
2836 CCAACTCGGAGAAATCCGTTTCAACAGCGGTGACGGTAACGGCAATT 2885  
||| .....  
877 .....IleLysLeuGlySerLeuLeuValAsnGlySerLe 888  
2886 GAACGCTCAGGGAACATTCGCTTATATCGAACTTCGCGTACCGCA 2935  
||| .....  
888 uSerLeuThrGlyGluAsnAlaAspIle.....L 898  
2936 GCGCAATTAAGCTGCGGAAAGTTCC..... 2964  
|||||.....  
898 ySgLyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArg 914  
2965 .....GAAGCACTTACACCTTGCGTCTGCACACATTCGCG 2999  
|||||.....  
915 AspThrLeuAsnIleThrGlyAsnPheThr.....AsnAsnG 927  
3000 CAACGAAACCCGTAAGTCTCGACAAATTGACGTAAGTGAAGAAAGCA 3049  
|||||.....  
927 yThrAlaGluIleAsnIleThrGlnGlyValValLysLeuGly.....A 942  
3050 ACACACCGCTGCGGAAATCTTAATTCACCTCGCAAAAGCAACAGCTC 3099  
|||||.....  
942 snValThrAsnAspGlyAspLeuAsnIleThrThrHisAlaLysArgAsn 958  
3100 GATGCGGCGCATGCGCTTATCAGCTTATCCGAAAGACGCGAGTCCG 3149  
|||||.....  
958 GlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGlySerLeuAs 975  
3150 CCTGCATATTCGCTCAAGAAAGACAGCTT..... 3180  
||| .....  
975 nIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyLysAsnIle 992  
3181 .....TCGCAAAACCTC..... 3192  
992 ergLInLysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr 1008  
3193 .....GGCAAGCGGAGAAACAGAGCGCGCTT 3221  
|||||.....  
1009 LysGlnIleThrIleLysLysGlyLysAspGlyLysAspSerSerSerAs 1025  
3222 GACGCAAAACAGCAACATTCGCGCAACACAGCGGCAAAAGACA 3271  
|||||.....  
1025 palAThrSerAsnAlaAsnLeuThrIleLysThrLysGluLeuLys... 1040  
3272 ACGCGAAACCTTGACGCGCTGATTTGGCGCGGCGCATGCCACCGAA 3321  
||| .....  
1041 .....LeuThrGluAspLeuSerIleSerGlyPheAsn..... 1051  
3322 AAGCGAAGAAAGTGTGCGCAACCGCGCGCGAG.....GCAGCGGCGCA 3365  
|||||.....  
1052 LysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsnSe 1068  
3366 AATGCGCGCATTTATGACAGCGGAGAGAGAAAAACGCGGCGAGGGG 3415  
|||||.....  
1068 rAsnAspGlyAsnSerGlyAlaGluAlaLysThrValThrPheAsnAsnV 1085  
3416 ATAAAGACACGCGCTTGCGAAACAG.....CGCGAAGC 3450  
|||||.....  
1085 alLysAspSerLysIleSerAlaAspGlyHisAsnValThrLeuAsnSer 1101  
3451 GAAACCGCGCGCTACACACGCGCTTCCCGCGCGCGCGCGCGCGCG 3500  
|||||.....  
1102 LysValLysThrSerSerSerAsnGlyArgGluSerAsnSerAspAs 1118  
3501 GGAATTGCGCGCAACCGCACCGCCCAACCGCAACCGCGCGCGCGCG 3550  
|||||.....  
1118 nasPThrGlyLeuThrIleThrAlaLysAsnValGluValAsnLysAspI 1135  
3551 TGATCAGC...CGTTATGCCAATAGCGGTTTGAGTAATTTCCGCAAG 3597  
||| .....  
1135 leThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr 1151  
3598 CTCACAGCGTTTTCGCGCTACAGACGAATTGACCGGCTTTGGCGA 3647  
|||||.....  
1152 ThrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrT 1168  
3648 AGACCGCGCGACGCGCTTGACAAAGCGGCAATTCGCGCACCAAACT 3697  
||| .....  
1168 hrLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerVal 1184  
3698 ACCGTTGCG...AAGATTCCGCGCTACCGCAACAAACCGACTGCGC 3744  
|||||.....  
1185 SerAlaThrGlyAspLeuThrThrLysSerGlySerLysIleGluAlaLys 1201  
3745 CAATCGGTATGCAAAAACTCGGACGCGGCGCGCTCGCATCTGTT 3794  
|||||.....  
1201 sSerGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly... 1216  
3795 TTCGCAACACCGGACCGGAAACCTTCGACAGCGGCAATCGGCA... 3838  
||| .....  
1217 .....ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1231  
3839 .....ACTGCGACGCGCTTGCGCAC 3858  
||| .....  
1232 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThr 1248

seq\_name: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1992.DAT:AA26503

ID	standard; 911 AA.
AA026503	standard; 911 AA.

DT 12-MAR-1993 (first entry)

atrophic rhinitis; alternative cleavage.  
B. bronchiseptica; P.68; outer membrane protein; piglet; probe;  
KW  
KW  
atrophic rhinitis; alternative cleavage.  
vv

05 Bordetella bronchiseptica.

Key Protein	Location/Qualifiers
FH	35..632
FT	

ET  
ET  
Dof

FT Pentido

PN XX W0921/587-A.

21-MAR-1992

XX  
XX  
XX

DR N-PSDB: AAQ3

PT atrophic rhinitis

CC The sequence

CC WHICH IS ASS

The sequence given is the P.94 antigen from *B. bronchiseptica*. The P.68 antigen is formed by alternative cleavage of this protein. P.68 is an outer membrane protein with a molecular weight of 68 kD which is associated with protection of piglets against atrophic

SQ Sequence 911 AA.

Quality: 279.00

97:233560 + HJ 001:

02 03 200 210D 003 X AAKZ03

1306 GCGCGGGCGTTCATATCA

**1356**

.....KTSY 71

```

... ..
82 LuAsnProAlaIaGlu..

```

[illegible]

1506 TGAATCGGCTTGGTCAGC

1556 ATCAGTTCACCCCGACCA

1000

121 valAlasprhIsAlatThrL

```

137  pg]vt]ea]at]t]vrv]
      :|||      :::

```

|||  
::: |||

1747 AAC TTGGATAGCAAAAAGG

1707 8888888888888888

.....Asp..... 100

191 eugInProGIuasPLeuPro

[illegible]

1935 ACCGACACCGCACGGCCTAC

```
1985 AAGGTATCCCAAGAGAAATCGTGTGGACAACAGATTGATCGACCGC 2034
    :|||
231 spgLy..... 232
2035 ACATTAAAGCGGAAAACTTCATATTCAGGCGGACACGCGTGTTC 2084
    ||||| ||||| :|||
233 .....GlyHisIleThrGlyIleArgAlaAla..... 241
2085 CCGCAAGTTGGCCAAAGTGGAGGCGGATTGGCATTTAAGCAATCACGCC 2134
    :||| ||||| :|||
242 .....GlyValAlaAlaMetaspLy..... 248
2135 AAGCAGTTTCGTGTGCGACCGCATCAAAAGCACACATCTGTACAGT 2184
248 ..... 248
2185 TCGGACTGGACGGGTGTGACAAATTGTACGAAAAAACATTACCGACGA 2234
248 ..... 248
2235 TAAAGTGTCTTCATTGACACAGACCGACATCAGAGCATGTCCAGC 2284
    :||| :||| :|||
249 .....AlaIleValHisLeuGlnArgAlaThrIleArg..... 259
2285 TTGCCGATCAGCTCATTTAATCTCACAGACTTGCACACTCAACGCGC 2334
260 .....ArgGly 261
2335 AATCTTAGTCAGGCGGAGACACGCACTATACGTTACGCGCAACGCCAC 2384
    :||| :||| :|||
262 AspAlaProIleGlyGly.....AlaValProIleGlyAlaVala 274
2385 CCAAAACGCG.....A 2395
274 IProIleGlyPheGlyProLeuLeuAspGlyTrpTyrGlyAlaAspValS 291
2396 ACCTCAGCCTCGTGGCAATGCCCAAGCACATTTAATCAAGCACATTA 2445
    ||| ||| ||||| :|||
291 erAspSerThrValAspLeuAlaGlnSerIleValGlnAlaProIleu 307
2446 AAGCGCAACATCGGCTTCGACATGCTTCATTTAATCAAC..... 2490
    :||| :||| :|||
308 GlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrValSerGlyG 324
2491 .....AACAAACGCGTACAAAACGCG..... 2511
324 ySerLeuSerAlaProHisGlyAsnValIleGlnThrGlyGlyAla 341
2512 .....AGTGTAGCGCTTTCGACACAC 2532
341 rGArgPheProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357
2533 GCTAGGCAAAAC.....GT 2546
    ||||| :|||
358 AlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuProGlnProva 374
2547 AAGCATTCGCGCACTCAAGCGCAATGTCTCCCTAGCCGATAAGGCAAT 2596
    :||| :||| :|||
374 lLysLeuThrLeuAlaGlyGlyAlaGlnGlyGlnGlyAsp.....IleV 389
2597 TCCATTTGAAAACACGCGCTTACCAGAAAATACAGCGCGGCGACAGAT 2646
    ||| :||| :|||
389 aAlaIleThrLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405
2647 ACGCATTCACCTTAAAGACAGCAATGAGACGTCGCGTGGCGACACGA 2696
    :||| :||| :|||
406 ValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaIleThrArgAl 420
2697 ATTAGCATTTAAACCTTGACACGCCACCATTTACACTCAATTCGCGCT 2746
    :||| :||| :|||
420 aValAlaSerLeuSerIleAspAlaAlaThrTrpValMetThrAsp.... 435
2747 ATGCACAGATCGCGCAGCGCGCAACCGCGCATGTGGCGCAGATCGCGC 2796
    :||| :||| :|||
436 .....AsnSerAsnValGlyAlaLeu 442
2797 GCGCGCGCTTTCGCGCGGTTCCTATTATTCGCTTACCGCCCAACTTCGC 2846
    ||| ||| :|||
443 ArgLeuAlaSerAspLySer...ValAspPheGlnGlnProAlaGlnAl 458
2847 AGAATCCGCTTTCACACGCTGACGCGTAACCGCAATTTGAACGTCAGG 2896
    :||| :||| :|||
458 a...GlyArgPheIysCysLeuMetValAsp...ThrLeuAlaGlySerG 473
2897 GAACATTCGCTTATGTGGAACCTTCGCTACCGGACGCGCAATTTG 2946
    ||| ||||| :|||
473 lLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLyLeu 489
2947 AAGTGGCGGAAAGTTCGGAAGCACTTCACCTTCAGCTGTCAACAATAC 2996
    :||| :||| :|||
490 ValValMetArgAspAlaSerIleGlnHisArgLeuLeuValArgAsnSe 506
2997 CCGCAACGAACCCGTAAAGTCTCGACAAATTGACGTAAGTGAAGGAANG 3046
    :||| :||| :|||
506 rGlySerGlnProAlaSerGlyAsnThrMetLeuLeuValGln..... 520
3047 ACAACACACCCCTGCGCAAAATCTTAATTTCACCTCGAAACGAA... 3093
    ||||| :||| :|||
521 .....ThrProArgGlySerAlaAlaThrPheThrLeuAlaSerAsp 535
3094 ...CACGTGATGCGCGCATGAGCGTATACGCTTATCCGCAAA...GA 3137
    :||| :||| :|||
536 GlyLysValAspIleGlyThrTyrArgTyrArgLeuAlaAlaSerGlyAs 552
3138 CCGCGAGTTCGCGCTGATATTCGCTCAACGAACAGCTTTCGACGA 3187
    :||| :||| :|||
552 nGlyGlnTrpSerLeu.Val..... 558
3188 AACTCGCAAGGCGGGAAGAACAGAGCGCGCTTGACGCAAAACAGCGA 3237
558 ..... 558
3238 CAACCTTCCGCGCAACACAGCGGGAAGAACACAGCGCGCAACGCTTGA 3287
558 ..... 558
3288 CCGCGCTGATTCGCGCGCGCGCAATGCCACCGCAAAAGGCAAGCTTG 3337
559 .....GlyAlaLysAlaProProAlaProLysProAla 569
3338 CCGAACCGCGCGCGAGCGAGCGGGAAGATCCGCGCATTTAGCAGCGC 3387
    ||| :||| :|||
570 Pro...GlnProGlyProGln..... 575
3388 GAGGAAGAGAAAAACGGGTGACAGCGGATTAAGACACCGCCTTGCGAA 3437
575 ..... 575
3438 ACACGCGGAAGCGGAAACCGCGCGGCTACACCGCTTCCCGCGCGCC 3487
576 .....ProGlyProGlnProProGln...ProProGlnTr 586
3488 GCCCGCGCGCGCGGATTTGCCCAACCGCAGCGCCCAACCGCAACCCAA 3537
    ||||| :||| :|||
586 oProGlnProProGlnArgGlnProGlnAla...ProAlaProGlnProPro 602
3538 CCGCAGCGGCACTGATCAGCGCTTATGCCAATAGCGGTTTGAGTGAATT 3587
    :||| :||| :|||
603 AlaGlyArgGlnLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrG 618
3588 TTCCGCGCAGCTCAACAGCGTT.....TTCCGCGTACAGAGCAATTCG 3631
    :||| :||| :|||
618 yGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGlnSerAsnAlaLeuS 635
3632 ACCGCGTGTTCGCGAAGACCGCGCAAC.....GCCGTTTGG 3669
    :||| :||| :|||
```







708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPhelLysLeuLysP 724  
1874 GCGGACAAATTTAAAC.....GGC 1893  
724 roAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740  
1894 AATATACGCAACA...AAGGCAACGTGTTTTCAGCGGACAGCCGAC 1940  
741 AsnIleThrAlaThrGlyGlySerValPhePheAsp..... 753  
1941 ACCGACGCTAC...AATCATTTAGAGACGGGTGTCAAAATGGAAG 1987  
754 .....IleTyAlaAsnHisSerGlyArgGly.....AlaG 764  
1988 GTATCCCAAGAGAAATCGTGTGGACAACGATTGGATGACCGCACA 2037  
764 LuLeuLysMetSerGluIleAsnIleSerAsn..... 774  
2038 TTTAAACGGAAAACTTC.....CATATTGAGGGCGGACAGC 2075  
775 .....GlyAlaAsnPhetThrLeuAsnSerHisValArgGlyAspAspAl 789  
2076 GGTGGTTCCCGCAATGTTCGCAAGAGGAGGCGATGGCATTTAAGCA 2125  
789 aPhe.....LysIleAsnLysAspLeuThrIleAsn. 799  
2126 ATCAGCGCCAAAGCAGTTTCGTGTGCGACCCGATCAAGCAACACATC 2175  
799 ..... 799  
2176 TGTACAGCTTCGAGCTGAGCGGTGTGCAAGTTTACGAAAAAACAT 2225  
800 AlaThrAsnSerAsnPh..... 805  
2226 TACCGAGATTAAGTGATGTCTTCATGACCAAGCCGACATGAGAGCA 2275  
806 .....SerLeuArgGlnThrLysAspSerPheT 815  
2276 ATGTCAGCTTCGCGATCAGCTCATTTAATATCATACAGAGCTTGCACA 2325  
815 TyrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831  
2326 CTCACAGCAATCTTACGAGCGGAGACACGCACTATACGGTTACGCG 2375  
832 LeuGlyLysValThrLeuGlyGly..... 840  
2376 CAAGCGCAACCAACGCAACTCAGCCCTGCGGCAATGCCCAAGCAA 2425  
841 .....GlnAsnSerSerSerIleThrGlyAsn.....IleT 852  
2426 CATTATTAACAGCC.....ACATTAAAGGCAACACATCGGCTCG 2466  
852 hTrIleGluLysAlaIleAsnValThrLeuGluAlaAsnAlaProAsn 868  
2467 GACAAATCTTCATTATTAATCTAACCAACACGCGTCAAAACGCGAGTCT 2516  
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerIle 882  
2517 GAGCGTTCCGACACGCTAAGCAACGTAAGCCATTCGCGACTCAAGC 2566  
882 uLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysG 899  
2567 GCAATGTCTCCCTAGCGATAGCAGTATTCAT.....TTTGA 2607  
899 LysAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915  
2608 AACAGCGCTTTACCGGAAAAATCAGCGGCGGCAAGATACGCGATTACA 2657  
916 ThrLeuAsnIleThrGlyAsnPhetThrAsnAsnGlyThrAlaGluIleAs 932  
2658 CTTAAAGACGGAATGACGAGCGTGGCGTGGGACGAGATTTGCAATT 2707  
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949

2708 TAAACCTTGACAACGCCACCATTAACCTCAATTCGCCCTATGACACGAT 2757  
949 euAsn.....IleThrThrHisAla..... 955  
2758 GCGGACGCGGCCAAACCGGCAAGTCGCGGAGATGCGCGCGCCGCTTC 2807  
956 .....LysArgAs 958  
2808 GCGCGCTTCCCTATTATCGGTTACCGCGCAACTTCGCGAGATGCCGTT 2857  
958 nGlnArgSerIle..... 962  
2858 TCAACACGCTGACGGTAAACGCAATTTGACGAGGAAACATTCCGC 2907  
963 .....IleGlyGlyAspIleIleAsn..... 969  
2908 TTTATGTCGAACTCTCGGCTACGCGGCGGCAAAATTTGAACTGGCGA 2957  
970 .....LysLysGlySerLeuAsnIleThrAs 978  
2958 AAGTTC.....G 2965  
978 pSerAsnAsnAspAlaGluIleGlnIleGlyLysAsnIleSerGlnLysG 995  
2966 AAGGCATTAACCTTGCT.....GTCAACATATACGCAACGAA 3006  
995 LuGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr..... 1008  
3007 CCCGTAAGTCTCGAGCAATTTGACGGTA.....GTGGAAGAAAGA 3047  
1009 .....LysGlnIleThrIleLysLysGlyIleAspGlyLys 1021  
3048 C...AACACACGCTGTCCGAAATCTTAATTTACCTGCAAAACGAC 3094  
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrLys..... 1035  
3095 ACGTGATCCGCGCATGCGCTTATCAGCTTATCCGAAAGACGCGAG 3144  
1035 ..... 1035  
3145 TTCGCGCTGCATTAATCCGCTCAAGACAGACGTTTCGCAAACTCG 3194  
1036 .....ThrLysGluLeuLysLeuThrGluAspLeu 1046  
3195 CAAGCGGAGAAACAGAGCGCCCTTGACGCGCAACAGGACACTTG 3244  
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063  
3245 CCGCCAAACACAGGCGGAAAAAGAACACACGCGCAAGC..... 3282  
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079  
3283 .....CTTGACGCGCTGATTGCGGCC...GGCGGCA 3311  
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHis 1096  
3312 TCCACCGGAAAGGCAAGAGTGTGCGAACCGCGCGGACGACGCG 3361  
1096 nValThr.....LeuAsnSerLysValLysThrSerSerAsnGlyG 1111  
3362 GCGAAATCCCGCATTTATGACGCGGAGGAGAAAGAAAGCGGTGCG 3411  
1111 LysArgGluSer.....Asn 1115  
3412 GCGGATTAAGACACGCGCTTG.....GCGAAACAGCGGAGCGGA 3452  
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131  
3453 AACCGGCGCGGTACACCGCCTTCCCGCGCGCGCGCGCGCGCGG 3502  
1131 ..... 1131

```

3503 ATTTGCCGCAACCGACCCCAACCGCAACCCACCGCGGCACTG 3552
1132 .....:||||:
3553 ATCAGC...CGTATGCCAATAGGGTTTGATGATTTTCCGCACGCT 3599
1136 ThrSerLeuValThrValAsnLeuThrAlaSerGluValThrThr.T 1152
3600 CAACAGCGTTTCCGCTACAGAGCAATTTGACGGCTGTTCCCGAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyValSerIleThrThr 1168
3650 ACCGCCGACAGCGCGTTTGACAAAGCGCATCGGGACACCAACTAC 3699
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerVal 1185
3700 GGTTCGC...AAGATTCCGGCGCTACCGCAACAAACCGACCTGGCA 3746
1185 rAlaThrValAspLeuThrThrLysSerGlySerIleGluAlaLysS 1202
3747 AATCGGTATCGCAAAAACCTCGCGACGGCGCGCATCTGTTT 3796
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216
3797 CGCACACCGGACCGGAAACCTTCGACGACGGCATCGGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839 .....ACTCGCGACGGCGTTGGCCACGG 3860
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249
3861 TGCCGTTTCCGGCAAT.....ACGGCATCGCGAGTTCGACATCGGCA 3904
1249 eutThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265
3905 TCAGCGCGCGCGGCGTTTGTAGTACGGCGACCTTCACAGCATCAGA 3954
1266 SerThrLysGlyGlnValAspLeuAla.....GlnAsnGlySerI 1280
3955 GGCAGAAATCGCGCGCGCTGCATTAAGCGCATTCAGCAAGATACCG 4004
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297
4005 CGCAGGTTTCGCGGATTCGGCATCGAACCAGCGCAACGGCT 4054
1297 eutThrThrValAlaGlySerAspIleLysAlaThrThrSerGlyThrLeuVal 1313
4055 ATTTCGTCGCAAA 4066
1314 IleAsnAlaLys 1317

seq_name: /SIDSL/gcgdata/geneseq/gene-seq-emb1/AA1992.DAT:AA25578
seq_documentation_block:
ID AAR25578 standard; Protein; 922 AA.
XX
AC AAR25578;
XX
XX 08-JAN-1993 (first entry)
XX
DE Bordetella parapertussis P95 antigen precursor.
XX
KM Whooping cough; P70 antigen; P95 precursor protein; vaccination.
XX
OS Bordetella parapertussis.
XX
XX
FH Key Location/Qualifiers
FT Protein 35..643
FT Binding-site /label= P70
FT 260..262
FT /note="motif associated with cell-cell adhesion"
FT Region 266..285

```

```

FT /note="contains 5 direct, tandem repeats"
FT Region 575..612
FT /note="contains 9 direct repeats of Pro-Gln-Pro"
FT Binding-site 712..714
FT /note="motif associated with cell-cell adhesion"
PN W09211292-A.
PD 09-JUL-1992.
PF 23-DEC-1991; 91WO-GB02302.
PR 21-DEC-1990; 90GB-0027901.
PA (WELL ) WELLCOME FOUND LTD.
XX Charles IG.
XX WPI; 1992-250033/30.
XX DR N-PSDB; AAO26509.
XX
XX Acellular vaccine for immunisation against whooping cough -
XX comprises protein uncontaminated by B. parapertussis components
XX and capable of binding antibodies which bind native P70 antigen
XX
XX Claim 1; Fig 1; 20pp; English.
XX
XX A cosmid library was constructed by transforming E.coli HB101 with
XX recombinant cosmids prepared by partial digestion of B. parapertussis
XX chromosomal DNA with Sma3A and cloning of 40-50kb fragments into the
XX CC BamHI site of cosmid pHC79. The cosmids were screened with a 1.8kb
XX CC ClaI fragment from the prn gene of B. pertussis. The insert from one
XX CC positive colony, harbouring cosmid pBD811, was sequenced and found to
XX contain an open reading frame encoding a 922 amino acid protein
XX with calculated mol.wt 95,177. This precursor protein ("p95") is
XX processed in vivo to the P70 antigen of apparent mol. wt. 70,000 as
XX determined by SDS-PAGE, but with actual mol.wt. 61kD. Antigenic
XX fragments of the protein will be useful in developing an acellular
XX CC vaccine against B. parapertussis. Preferred fragments include amino
XX acids Pro577 to Pro612 or Ala574 to Pro612.
XX
SQ Sequence 922 AA:

alignment_scores:
Quality: 275.50 Length: 1162
Ratio: 0.566 Gaps: 54
Percent Similarity: 41.910 Percent Identity: 21.687

alignment_block:
US-09-303-518D-653 x AAR25578 ..

Align seg 1/1 to: AAR25578 from: 1 to: 922

1366 AACGACCGCCTGTCAAAATCGGCAAAAGC.....ACGCT 1400
||| |||||:|||| |||||
2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThrTh 18
1401 GCTGTTCAAGCCAAAGG..... 1419
||||:||||| |||
18 rIleuAlaMetAlaLeuGlyAlaLeuGlyAlaAlaProAlaAlaTYrAlaA 35
1420 .....GAAGCAAGAGCTCGTACGCGGCGAGGATAA...GTATC 1461
35 sPTTPhaAsnGlnSerIleIleLysAlaGlyGluArgGlnHisGlyLe 51
1462 TTATGATCAGCAGCGCGGATCATGACGCAAAACAGCTTATGTAAT 1511
||||:||||| ||| ||| ||| |||
52 HisIleLysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrTh 68
1512 CGGCTTGTCAGCGCGGCGGAGCGTGCATCACTGAATGCCATTAATCACT 1561
||||| ||||| ||| ||| ||| |||
68 rIleLysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn.... 83

```

1562 TCACCCGCAAAAGCTATTTCCGCTTCCGCGCGACGTTGGATTG 1611  
84 .....ProAlaIaGluLeuArgPheGlnAsnGlySerValThrSer 97  
1612 AACGGCATTCGCTTCTGTTCCACCGCATTTCAAAATACCGATGAAGGGC 1661  
98 SerGly.....GlnLeuPheAspGluGlyVal 106  
1662 GATGATTGTCCACCAATCAAGACAAAGAAATCCACGCTTACATTACAG 1711  
106 1.....ArgArgPheLeuGlyThrValThrValLysSA 117  
1712 GCATATAA.....GATTATTACTACACCGCGCATTAACACACTTG 1752  
117 1aGlyIysLeuValAlaAspHisAlaThrLeuAlaAsnValSerAspThr 1799  
1753 GATAGCAAAAAAGAAATTCGC...TACAACGGTTGGTTGGCGAGAAAGA 1799  
134 ArgAspAspAspGlyLeuAlaLeuThrValAla.....GlyGluGln 147  
1800 TGCACCAAAACGACACGCGCGCTCAATCTGAATTACCACGGAGAGAG 1849  
148 .....AlaGlnAlaSerIleA 153  
1850 CGGATCGACATTACTGCTTCCGCGCGAGACAATTTA.....AACGC 1893  
153 1aAspSerThrLeuGlnGlyAlaGlyValArgValGlnArgGlyValA 169  
1894 AATATCAGCCAAACAAACGCAACTGTTTTCAGCGCGACACCGACACC 1943  
170 AsnValThrValGlnArgSerThrIleValAspGlyGlyLeu 183  
1944 GCACGGCTCAATCAATTAGAGAGCGGCTGTCA...AAATGAGAGTA 1990  
184 .....HisIleGlyThrLeuGlnProLeuGlnProGluAsp 196  
1991 TCCCAAGAGAAATCTGTGGGACACGATTGATGACCGC..... 2034  
196 eupProSerArgValValLeuGlyAspThrSerValThrAlaValPro 212  
2035 .....ACATTTAAAGCGGAAACCTTCCA 2057  
213 1aAspGlyAlaProAlaAlaValPheValPheGlyAlaAsnGluLeuThr 229  
2058 TATTCAGCGCGACAAAGCGGTGTTCCCGCAATGTTGCCAAAGTGAAG 2107  
229 rValAspGlyGly..... 233  
2108 GCGATTGGCATTTAAGCAATCAGCGCCACAGAGTTTGGTGTGCGACCG 2157  
234 .....HisIleThrGlyArgAlaAla.....GlyValAla 244  
2158 CATCAAGCGACACAAATCTGTACACGTTCCGACTGACGCGGTCTGACAG 2207  
244 ..... 244  
2208 TTGTACCGAAAAACATTACCGAGCATTAAGTATGATCTTCAATTGACA 2257  
245 .....AlaMetAspGlyAlaIleValHisLeuGlnA 255  
2258 AGACCGACATCGAGAGCAATGTCAAGCTTGGCCATCAAGCTCAATTAAT 2307  
255 rGlnThrIleArg..... 259  
2308 CTCACAGACTTGCACACTCAAGCGCAATCTTAGTCAGGCGAGACAC 2357  
260 .....ArgGlyAspAlaProAlaGlyGly..... 267  
2358 GCATCTATACGGTTACGCGCAAGCGCCACCAAAACGCGCAACTC..... 2400  
268 .....AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282



```

FT      /label= repeat
FT      599..601
FT      /label= repeat
XX      W09115571-A.
XX      PD
XX      17-OCT-1991.
XX      PF
XX      28-MAR-1991; 91WO-GB00487.
XX      PR
XX      02-APR-1990; 90GB-0007416.
XX      (WELL ) WELLCOME FOUNDATION LTD.
XX      PA
XX      Clare JJ, Romanos MA;
XX      PI
XX      WPI: 1991-325214/44.
XX      DR
XX      N-PSDB: AA014319.
XX      PT
XX      Pichia microorganism transformants - for production of
XX      Bordetella pertactin antigens for whooping cough vaccines
XX      PS
XX      Disclosure; Fig 1B; 38pp; English.
XX      CC
XX      Pichia microorganisms are transformed for the expression of
XX      pertactin antigens. DNA sequence used are represented in AA014319-20
XX      encoding the B. bronchiseptica P.68 and B. pertussis P.70 antigen
XX      respectively or the B. pertussis P.69 encoding sequence described
XX      by I.G. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80:3554-3448
XX      (1989).
XX      SO
XX      Sequence 911 AA;

alignment_scores:
    Quality: 274.00      Length: 1149
    Ratio: 0.584        Gaps: 50
    Percent Similarity: 40.818      Percent Identity: 20.540

alignment_block:
US-09-303-518D-653 x AAR14320 ..

Align seg 1/1 to: AAR14320 from: 1 to: 911

1306 GCGCGGGCGTTTCATATCATGATGCGACAGTACCGTTACTTGAAAATGAAA 1355
|||||
58 GlyAlaGlyValArgThrAlaThrGlyThrThrIle.....LysValSer 72

1356 CGGCGTGGCAACGACGCCCTGTCCAAATGCGCAAGCAGCGTGTGG 1405
|||||
72 rGly.....ArgGlnAlaGlnGlyValLeuLeuG 82

1406 TTCACAGCCAAAGGGAACCAAGAGCGTGTGACAGGTGGGAGGAGTAA 1455
|||||
82 LuAsnProAlaAlaGln.....LeuArgPheGlnAsnGlySer 94

1456 GTCAATCTTAGTACAGCAGCGGAGATCAGCAAAAAACAAGCCTTAG 1505
|||||
95 ValThrSerSerGlyGlnLeuPheAspGlnGlyValArgArgPheLeu.. 110

1506 TGAATATCGCTTGGTCAAGCGGAGGAGGAGCGTGCACAACTGAATCCGATA 1555
|||||
111 .....GlyThrValThrValLysAla.... 117

1556 ATCAGTTCACCCCGACAAACTTATTTGCGCTTCCGGGCGAGCGTTTG 1605
|||||
118 .....GlyLysLeu 120

1606 GATTGAAAGGCGATTCGCTTCTTCACCGCATTAATAATACGATGA 1655
|||||
121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgAspAspAs 137

1656 AGGGGCGATGTGTCAACGACATCAAGCAAAAGATCCACGCTT.... 1701

```

```

137 pGlyIleAlaLeuTyrValAlaGlyGlnGlnAlaGlnAlaSerIleAla 154
|||||
1702 .....ACCATTCAGCCAAATTAAGATATTTACTACACCGCGCAATACAC 1746
|||||
154 spSerThrLeuGlnGlyAlaGlyValArgValAlaGlyAlaAsn 170
1747 AACTTGGATAGCAAAAAAGAAATTTGCTACACGCTTGTGGCGAGAA 1796
|||||
171 ValThrValGlnArg..... 175

1797 AGATGCAACCCAAACGAAGCGCGCTCATCTGCAAT.....T 1834
|||||
176 ....SerThrIleValAlaSpGlyLeuHisIleGlyThrLeuGlnProL 191

1835 ACCAAGCGGAAGAGCGCATTCGCATTTACTGCTTTCGCGGCAACAAAT 1884
|||||
191 eugInProGlnAspLeuProProSerArgValAlaLeuGlyAspThrSer 207

1885 TTAAACGGCAATATCACGCCAAACAGCGCAAACTGTTTTCAGCGGACG 1934
|||||
208 ValThrAla.....ValProAlaSerGly..... 215

1935 ACCGACACCGGACGCGCTTACATCATTTAGGAGCGGCGTCAAAANTG 1984
|||||
216 ....AlaProAlaAlaValSerValPheGlyAlaAsnGlnLeuThrValA 231

1985 AAGGTATCCCAAGAGAAATCGTGTGGCAACAGATTGGATGACCGC 2034
|||||
231 spGly..... 232

2035 ACATTTAAGCGGAAACTTCATATTCAGCGCGGAGCGGTGTTTC 2084
|||||
233 .....GlyHisIleThrGlyGlyArgAlaAla..... 241

2085 CCGCATGTTGCCAAGATGAGAGCGGATTTGCAATTAAGCAATCAGCGCC 2134
|||||
242 ....GlyValAlaAlaMetAspGly..... 248

2135 AAGCATTTTCGCTGTGCGACCGCATCAACCAACCAACATGTGTACAGT 2184
|||||
248 ..... 248

2185 TCGGACTGACAGGGGTGTGACAGTGTACCGGAAAAACCATTCAGCAGCA 2234
|||||
248 ..... 248

2235 TAAAGTATGTTCTTATGAGCAAGACGACATCAGAGCGCAATGTGAGCC 2284
|||||
249 ....AlaIleValHisLeuGlnArgAlaThrIleArg..... 259

2285 TTGCCCATTCAGCTCATTTAATCTCACAGGACTTCCACACTCAAGCGC 2334
|||||
260 .....ArgGly 261

2335 ATCTTAGTCAGCGGAGACAGCAGCATATACGTTACCGGCAAGCGCAC 2384
|||||
262 AspAlaProAlaGlyGly.....AlaValProGlyGlyAlaValA 274

2385 CCAAAACGCG.....A 2395
|||||
274 IProGlyGlyPheGlyProLeuLeuAspGlyTyrGlyValAlaSpValS 291

2396 ACCTCAGCGCTGTGGCGCAATGCCCAAGCAACATTTAATCAAGCCATTA 2445
|||||
291 etAspSerThrValAspLeuAlaGlnSerIleValGlnValProGlnLeu 307

2446 AACGGCAACACATCGCTTCGAGCAATGCTTCATTTAATCAATCAAGC.... 2490
|||||
308 GlyAlaAlaIleLeuArgAlaGlyArgGlyAlaArgValThrValSerGly 324

2491 .....AACACGCGCTACAAAACGCG..... 2511
|||||

```

```

324 ySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyAla 341
2512 ..... AGCTGACGCTTCCGCAAC 2532
341 rGaTgPheProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357
2533 GCTAAGGCAAC.....GT 2546
358 AlaArgAlaIleGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProVa 374
2547 AAGCCATTCCGCACTCACAGCGCAATGTCTCCCTAGCCGATAGGCAAT 2596
374 LysLeuThrLeuAlaGlyIleGlnGlyArgValLeuProGluProVa 389
2597 TCCATTGAAACAGCGCTTACCGGAAATTCAGCGCGGCAAGAT 2646
389 AlaIleArgGluLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405
2647 ACGGCAATTACCTTAAGACAGCGAATGAGCGCTGCGGCGACGCA 2696
406 ValAlaLeu...AlaSerGlnAlaArgTyrThr...GlyAlaThrArgAl 420
2697 ATTAGGCAATTTAACCTTGACACAGCGCAATTAACATTCGCGCT 2746
420 aValAspSerLeuSerIleAspAsnAlaThrTyrValMetThrAsp.... 435
2747 ATGCACACGATCGCGCAGCGCGCAACCGCGGCGGCGAGATGCGCG 2796
436 .....AsnSerAsnValGlyAlaLeu 442
2797 CGCGCGCTGCGCGCGCTTATATATTCGCTTACGCGCAATTCGCG 2846
443 ArgLeuAlaSerAspGlySer...ValAspPheGlnGlnProAlaGluAl 458
2847 AGAATCCCGTTTACACACGCTGACGCTAAACGGCAATTGACGCTAG 2896
458 a...GlyArgPheLysCysLeuMetValAsp...ThrLeuAlaGlySerG 473
2897 GAACATTCCGCTTATGTGCGAATCTTTCGCTACCGCGGCAATTCG 2946
473 LysLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspGlyLeu 489
2947 AAGCTGCGGAAAGTCCGAGGCACTTACACCTTGCCTGTCACAAATAC 2996
490 ValAlaMetArgAspAlaSerGlyGlnHisArgLeuValArgAsnSe 506
2997 CGGACAGCAACCCGTAAGTCTCGAGCAATGACGCTAGTGGAGCAAG 3046
506 rGlySerGluProAlaSerGlyAsnThrMetLeuValGln..... 520
3047 ACAACACACGCTGTCCGAAATCTTAATTTCACCTGCAAAAGCA... 3093
521 .....ThrProArgGlySerAlaIleThrPheThrLeuAlaAsnLysAsp 535
3094 ...CACTGATGCCGCGCATGCGCTATACGCTTATCCGCAAA...GA 3137
536 GlyLysValAspIleGlyThrTyrArgTyrArgLeuAlaIleAsnGlyAs 552
3138 CGGGAGTTCGCGCTGCATATTCGGGTCAAAAGAACAGAGCTTCCGACA 3187
552 nGlyGlnTrpSerLeu.Val..... 558
3188 AACTCGGCAAGCGGAGAAACAGAGGCCGCTTGACGGCAAAACAGGCA 3237
558 ..... 558
3238 CAACTGCCGCCAAACAGAGCGGAAAGACACAGCGCAAGCCTTGA 3287
558 ..... 558
3288 CGCGGTGATTCGGCGCGGCGCAATGCCAGCAAAAGCGCAAGTGTG 3337
559 .....GlyAlaLysAlaProProAlaProLysProAla 569

```

```

3338 CGAAGCGCGCGCGGACAGCGGAGAAATGCCGCAATTATGACAGCG 3387
570 Pro...GlnProGlyProGln..... 575
3388 GAGGAAGAAAAACGGGTGCAGCGGATTAAGACACCGCTTGGCGAA 3437
575 ..... 575
3438 ACAGCGCGAAGCGGAACCGCGCGCTTACGACCGCTTCCCGCGCG 3487
576 .....ProGlyProGlnProProGln.ProProGlnTr 586
3488 GCGCGCGCGCGGATTTGCGCAACCGCGCGCGCGCAACCGCAAC 3537
586 ProGlnProProGlnArgGlnProGlnAla.ProAlaProGlnProPro 602
3538 CGGACCGCGCGCTGATGACCGCTTATGCAATAGCGGTTCGATGAAT 3587
603 AlaGlyArgGluLeu...SerAlaAlaAlaAlaAlaAlaAlaAlaThrG 618
3588 TTCGCGCGCTCACAGCGTT.....TTCGCGCTACGAGCAGCAATTCG 3631
618 yGlyValGlyLeuAlaSerThrLeuTyrTyrAlaGluSerAsnAlaLeuS 635
3632 ACCGCGTGTTCGCGAAGACCGCGCGCAAC.....GCCGTTTG 3669
635 eLysArgLeuGlyGlyLeuArgLeuAsnProAspAlaGlyGlyAlaTrp 651
3670 ACAAGCGCATCGGAGACCAACACCTACGCTTCGCAAGATTCGCGCG 3719
652 GlyArgGlyPhe.....Al 656
3720 CTACCGCGCAACAGCACTTCGCGCAATGCGTATGCGAATAAATCTCG 3769
656 aGlnArgGlnGlnLeuAsp..... 662
3770 GCAGCGCGCGCTGCGCATCTTTCGCGACACCGCGAGAAACACC 3819
663 .....AsnArgAlaGlyArgArg 668
3820 TTCGACGAC.....GGCATCGGCAATCGCGCG 3848
669 PheAspGlnLysValAlaIleGlyPheGluGlyAlaAspHisValAla 685
3849 GCTTGCAGCGGTGCC.....GTTTGCAGCAATACGCGCATCGCGAGT 3892
685 aValAlaGlyGlyArgTyrGlnLysGlyGlyLeuAlaGlyTyrThrArgG 702
3893 TCGACATCGGCATCAGCGCGCGCGGTTTATGAGCGCGAGCTT... 3939
702 LysArgGlyPheThrGlyAspGlyGlyLysThrAspSerValHis 718
3939 ..... 3939
719 ValGlyGlyTyrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspAl 735
3940 .....TCAG 3943
735 aThrLeuArgAlaSerArgLeuGlnAsnAspPheLysValAlaGlySerA 752
3944 ACGGC.....ATCAGAGCAAAATCCGCGCGCGCTGCGATTAAGCGC 3987
752 spLysTyrAlaValLysGlyLysTyrArgThr.....HisGly 764
3988 ATTACAGCAAGATACCGCGAGT.....TTCGCGCGAT 4022
765 ValGlyAlaSerLeuGlnAlaGlyArgTyrPheAlaHisAlaAspGlyTr 781
4023 CGGCATGCAACCGCACATCGCGCAACCGCTATTTCTGCAAAAGCGG 4072
781 PheLeuGluProGlnAlaGluLeuAlaValPheArgValGlyGlyLys 798

```

[illegible]





2344 .....GCAGCGGAGACGACGACTATACGGTTACGGCGCAACGCCAC 2384  
 ||||| : : : : :  
 711 AsnLeuylsGluGlyAlaIylValAsnPhelyleuLysProAsnGluAs 727  
 2385 CCAAAAC.....GGCACTCAGCC 2404  
 : : : : :  
 727 metAsnThrSerLysProLeuProIleArgPheLeuAlaSnIleThra 744  
 2405 TCGTGGCAATGCCCAACACATTTAAT...CAAGCCATTTAAGCGC 2451  
 : : : : :  
 744 IaThnGlyGlySerValPhePheAspIleTyraIaAsnHisSerGly 760  
 2452 AACACATCGCTTGGACATGCTTCATTTAATTAAGCAACAGCGCGT 2501  
 : : : : :  
 761 ArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAsnGlyAla.. 776  
 2502 ACAAACGGCAGTCTGACGCTTCCGACACGCTAAGCA..... 2541  
 : : : : :  
 777 .....AsnPhelrIleuAsnSerHisValArgGlyAspAspAlaP 790  
 2542 .....ACGTAGCCATTCGCGACTCAACGGCAATGCTCC 2577  
 : : : : :  
 790 helysIleAsnLysAspLeuThrIleAsnIaIaThrAsnSerAsnPheser 806  
 2578 CTACCCGATAAGCAGATTCATTTGAA..... 2607  
 ||| : : : : :  
 807 LeuArgIaIaThrLysAspAspPheTyraSpolyTyraIaArgAsnAlaI 823  
 2608 .AACAGCCGCTTACCGGAAAAATCAGCGCGGCAAGATACGGCATTC 2656  
 ||||| : : : : :  
 823 eaSnSerThrTyraSnIleSerIleLeuGlyLysnValThr...Leug 839  
 2657 ACTTAAAGACAGCAATGACGCTGCGTGGCGACGGAATTAGGAAT 2706  
 : : : : :  
 839 LylGlyGlnAsnSer.....SerSerSerIleThnGlyAsn 850  
 2707 TTTAACTTGAACAAC...GCCACATTACACTCAATTCGCGCTATCGACA 2753  
 : : : : :  
 851 IleThrIleGluLysAlaIaIaSnValThrLeu..... 861  
 2754 CGATGCGGAGCGCGCAACCGGCACTGGCGGAGATCGCGCGCGCC 2803  
 : : : : :  
 862 .....GluAlaSnAsnAlaProAsnGln 870  
 2804 GTTGGCGCGTTCCTTATTCCTTATTCGCTTACGCGCACTTGGCGAGATTC 2853  
 : : : : :  
 870 IaaSnIleIaIaGAspArgVal.....Ile 877  
 2854 CGTTTCAACAGCTGACGGAATGGAACGATTCAGGAGACATT 2903  
 : : : : :  
 878 LysLeuGlySerLeuValAsnGlySerLeuSerLeuThrGlyLys 894  
 2904 CCGCTTATGTCGGAACCTTTCGCTACCGCAGCGGCAATTAAGCTGC 2953  
 : : : : :  
 894 naIaAspIle.....LysGlyAsnLeuThrLies 904  
 2954 CGGAAAGTTC.....GAA 2967  
 : : : : :  
 904 erGluSerAlaIaThrPheLysGlyLysThrArgAspThrLeuAsnIleThr 920  
 2968 GGCATTACACCTTGGCTGTCAACATACGGCAACAGACCCGTAAGTCT 3017  
 : : : : :  
 921 GlyAsnPhelThr.....AsnAsnGlyThraIaGluIleAsnI 933  
 3018 CGACACATTCAGCTAGTGAAGAAAAAGACACACACCGCTGCCAAA 3067  
 : : : : :  
 933 erThnGlnGlyValIaLysLeuGly.....AsnValThrAsnAspIa 948  
 3068 ATCTTAATTCACCCCTGCAAAACGACACGCTGATCGCGGCGATGCGGT 3117  
 : : : : :  
 948 spleuAsnIleThrThrHisAlaLysArgAsnGlnArgSerIleIleGly 964  
 3118 TATACGCTTATCCCAAGACGCGGAGTTCCGCTGCATATCCGGTCA 3167  
 : : : : :  
 965 GlyAspIleIleAsnLysLysGlySerLeuAsnIleThraAspSerAsn 981  
 3168 AGAACAGAGCTT..... 3180  
 : : : : :  
 981 naSpAlaGluIleGlnIleGlyLysAsnIleSerGlnLysGluLysnL 998  
 3181 .....TCCGACAAATC..... 3192  
 : : : : :  
 998 eutThrIleSerSerAspLysIleAsnIleThrLysGlnIleThrIleLys 1014  
 3193 ...GGCAAGCGGAGAGAACAGAGCGGCTTGCAGCAACAGCGACA 3239  
 : : : : :  
 1015 LysGlyIleAspGlyLysPserSerSerAspAlaIaThrSerAsnIaAs 1031  
 3240 ACTTCCCGCCAAACACAGCGGAGAAAAAGCAACGCGCAAGCTTGACG 3289  
 : : : : :  
 1031 nLeuThrIleLysThrLysGluLeuLys.....LeuThrGlu 1044  
 3290 CGCTGATTCGCGCGCGCAATGCCACGAAAGCGAGAAAGTGTTCG 3339  
 : : : : :  
 1044 spleuSerIleSerGlyPheAsn.....LysAlaGluIleThrAla 1057  
 3340 GAACCGCGCCGCGAG.....GCAGCGGAGAAATCCGCGCATTTATGCA 3383  
 : : : : :  
 1058 LysAspGlyArgAspLeuThrIleGlyAsnSerAsnAspGlyAsnSerG 1074  
 3384 GCGGAGAGAGAAAGAAAAAGGCTGCAGCGCGATTAAGACACCGCTTG 3433  
 : : : : :  
 1074 yAlaGluAlaLysThrValThrPheAsnAsnValLysAspGlnLysIle 1091  
 3434 CGAAACAG.....CGGAGACGGAACCGCGCGCTAC 3468  
 : : : : :  
 1091 erAlaSpGlyHisAsnValThrLeuAsnSerLysValLysThrSerSer 1107  
 3469 ACCGCTTCCCGCGCGCGCGCGCGCGCGGATTCGCCCAACCGCA 3518  
 : : : : :  
 1108 SerAsnGlyLysArgLysSerLysSerAsnSerAspAsnAspThrGlyLeuThrI 1124  
 3519 GCCCAACCGCAACCCCAACCGCGGACCTGATCAGC...CGTTATG 3565  
 : : : : :  
 1124 erThrAlaLysAsnValGluValAsnLysAspIleThrSerLeuLysThrV 1141  
 3566 CCAATAGCGGTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTCCG 3615  
 : : : : :  
 1141 alAsnIleThrAlaSerGluLysValThrThr.ThraIaGlySerThrI 1157  
 3616 GTACAGACGAATTTGACCGCGGTGTTCGGAAGACCGCGACGCGCT 3665  
 : : : : :  
 1157 eaSnAlaIaThrAsnGlyLysAlaSerIleThrThrLysThrGlyAspLies 1174  
 3666 TTGGAACACCGCATCCGGACACCAACACATACCTTCCG...AAGATT 3712  
 : : : : :  
 1174 erGlyThrIleSerGlyAsnThrValSerValSerAlaThrGlyAspLeu 1190  
 3713 TCCGCGCTACGCGCAACCAACCGACCTGCGCAATCGGTATGCAAAA 3762  
 : : : : :  
 1191 ThrThrLysSerGlySerLysIleGluAlaLysSerGlyGluAlaAsn 1207  
 3763 AACCTGCGAGCGGCGCGCTGCGCATCTGTTTTCGACACACCGACCG 3812  
 : : : : :  
 1207 IThrSerAlaThrGlyThrIleGlyLys.....ThrIleSerGly 1221  
 3813 AATCCTTCGACGACGCGCATCGCA..... 3838  
 : : : : :  
 1221 snThrValAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGly 1237  
 3839 .....ACTGCGACGCGCTTCCGCGCGGCGCTTTCGCGGCA 3876  
 : : : : :  
 1238 AlaGluIleAsnAlaThrLysGluLysAlaIaThrLeuThrAlaThrGlyAs 1254  
 3877 T.....ACGGCATCGGAGGTTCCACATCGCATCGCGCGCGCGG 3920

```

1254 nThreuthrthrthruAlaIagIysSerIleThrSerThrIysGlyIlnV 1271
1271 TTTTATGTCGGGACGCTTTTCAGACGCGATAGAGCAAAATCCGCCGCC 3970
1271 ..... 1271 ..... 1271 ..... 1271 ..... 1271 .....
1271 AlaSplLeuAla.....GlnsngIysSerIleAlaGlySerIleasn 1285
3971 GCGTCTCATTTAGCGCATTCAGCAAGATACCGCGCGGCTTCGGCGGA 4020
1286 AlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValAlaGl 1302
4021 TTCGCGATCGAAGCGACATCGCGCAAGCGCTATTTCGTCGCAAA 4066
1302 ySerAspIleIysAlaThrSerGlyThrLeuValIleAsnAlaIys 1317
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: AAB01827
seq_documentation_block:
ID AAB01827 standard; Protein; 969 AA.
AC AAB01827;
XX
XX 11-SEP-2000 (first entry)
XX
XX Haemophilus influenzae strain Jcyc mature HMW2A protein, SEQ ID NO:32.
XX
XX Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX non-typable Haemophilus influenzae; NTHI; non-encapsulated;
XX recombinant production; Escherichia coli; anti-bacterial; vaccine;
XX human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.
XX
XX Haemophilus influenzae strain Jcyc.
XX
XX WO200020609-A2.
XX
XX 13-APR-2000.
XX
XX 07-OCT-1999; 99WC-CA00938.
XX
XX 07-OCT-1998; 98US-0167568.
XX 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX WPI; 2000-303789/26.
XX N-PSDB; AAA52178.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
XX PT proteins of Haemophilus which are used as a vaccine to provide
XX PT protection against Haemophilus induced diseases in humans -
XX
XX Claim 8; Fig 19A-O; 307pp; English.
XX
XX The invention relates to the recombinant production of Haemophilus
XX influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX expression construct used to effect recombinant expression comprises a
XX promoter functional in E. coli (e.g., the T7 promoter) operably linked
XX to a modified hmwABC operon from a non-typable (non-encapsulated) H.
XX influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene
XX clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
XX and the hmwB and hmwC genes encode accessory proteins which are
XX responsible for post-translational processing and secretion of the HMW
XX proteins. The modified hmwABC operon used in the expression construct of
XX the invention contains an A gene modified such that it encodes only the
XX mature HMW. The invention also discloses hmwA genes (AAA52175-A52188)
XX and hmwA proteins (AAB01824-B01849) from the non-typable H. influenzae
XX strains Jcyc, KI, K21, LCD2C, PMH1, 15 and 12. The nucleic acids and
XX vectors are used for the production of recombinant H. influenzae HMW
XX proteins which can be used as vaccines to mediate a humoral or

```

```

CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMW protein from a non-typable
CC strain of H. influenzae.
XX
XX Sequence 969 AA:
SQ
XX
XX alignment_scores:
XX Quality: 273.00 Length: 1164
XX Ratio: 0.487 Gaps: 54
XX Percent Similarity: 48.110 Percent Identity: 20.619
XX
XX alignment_block:
XX US-09-303-518D-653 x AAB01827 ..
XX
XX Align seg 1/1 to: AAB01827 from: 1 to: 969
XX
XX 1090 AATGTTTCTTTATTC.....GAGACGACAGAGAGACC 1121
XX |||||.....
XX 3 AsnValSerIleGlnAsnProSerThrGluArgAsnAspSerAsnGlns 19
XX 1122 TGTTTATCAGCTGCGAGTGGGTCAACAGTTATGCAACCGACGCTGAATA 1171
XX :|||:.....
XX 19 pleuGluTyrThrGlyThrGlyGlnAsnIleAsnAsnProIysValAsnA 36
XX 1172 ATGGA.....GAAATATAT..... 1185
XX |||||.....
XX 36 snGlnSerIysLysThrLeuThrSerSerIleLeuGlnAsnIleLeuLys 52
XX 1186 .....TCCTTATTTGAC...AAGGAAAGGTGAATTCATCTACCGAG 1226
XX |||||.....
XX 53 LysGlySerPheValAsnIleThrAlaThrAspAsnIleTyrValAsnSe 69
XX 1227 CAAATCAACCAAGGC...GCGGCGGTTTGTATTTTGAAGGTAATTTTA 1273
XX |||||.....
XX 69 rSerIleAsnIleGlyAspSerGlyHisLeuIleLeuSerGlyGlyA 86
XX 1274 CGGTCTCGCTTAATAAACAAACGAAACGTGGCAGCGCGCGCTTCATATC 1323
XX 86 rg.....AsnGlyGlyValLysIle 93
XX 1324 AGTATGCGCAGTACCGTTACTTGAAGAATACGCGCGTGGCAACGACCG 1373
XX :|||:.....
XX 94 AsnLysAsnIleThrSerThr..... 100
XX 1374 CTGTCCAAATATGCGCAAGCGCGCTGTGTTCAAGCAAGG..... 1419
XX 101 .....GlyGlySerLeuThrIleAsnSerLysGlyTyrP 112
XX 1420 ..GAAACCAAGGCTCGTACGCGTGGC..... 1446
XX :|||:.....
XX 112 alaSplIleIserAsnIleSerLeuGlyThrGlyPheLeuAsnIleThr 128
XX 1447 ...GACGGTAAAGTCATCTTAGTCAGACGCGGACGATCAAGCAAAA 1493
XX :|||:.....
XX 129 SerAsnIysSerValAlaPheGlyLysAlaAspLysAspLysAlaArgSe 145
XX 1494 ACAAGCCTTAGTGAATAGCGCTGTGTCAGCGCGGACGAGCGAGTCGCAAC 1543
XX :|||:.....
XX 145 rAlaAlaAspAlaGlnIle.....ValAlaGlnGlyIleIleAsn 159
XX 1544 TGAATGCGGATATTCAGTTCAACCCGACAAACTATATTCGCGCTTCGC 1593
XX |||||.....
XX 159 eutHrGlyGlnAsnLys.....ThrPheArg 167
XX 1594 GCGCGACGTTGATTTGAACGGCGCATTCGCTTCGTCACCGCATTA 1643
XX :|||:.....

```

```

168 LeuAsnValSerLeuAsnGlyValGlyGlnGlyLeuSerIleThrSe 184
1644 AATACCGATGAGAGGGCGATGATTGTCAACCAATCAAGCAAAAGAT 1693
184 rsn.....ValGlyAsnGlnIleThrIleHisIleHisPheAsp 195
1694 CCACCGTTCCATTTACAGCAATTAAGATATTACTACAAACCGC..... 1737
195 LylIleAsnIleThrGlyAsnValThrIleAsnGlnThrAlaPheAla 211
1738 .....ATAACAAACA 1748
212 ThrThrAlaIleTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSer 228
1749 CTTCGATACGAAAAAGAAATTCCTAC.....A 1777
228 rleuAsnValGlnIleAsnSerSerPheThrIleIleValGlyThrGluS 245
1778 ACCGTTGGTTGGCGAAGAAATGCAACCAACGACGGCGCTCAAT 1827
245 eAsnAlaIlePheGlyProThrThrProLeuArgSerSerGlyValPhe 261
1828 CTGAATTTACCAACGGAGAGGATGCACTTACTGCTTCGCGCG 1877
262 PheAsn.....G 264
1878 AACAAATTTAAACGGCAATATC.....ACGCAACAAACGGCAAC 1918
264 YThr.....AsnGlyAsnMetValLeuAsnValGlyThrAsnSerArgV 279
1919 TGTTTTACAGGCGGACGACCGACCGACCGCTTCAATCATTTAGGAAGC 1968
279 AlLeuPheAsnLeuIleAsnProAsnGlnAsnIleThrAlaIleGly 291
1969 GGGTGGTCAAAAATGAGATATCCACAGAGAAATCGTGGGACAA 2018
292 .....SerLysPro..... 294
2019 CGATTGGATGACCGCACCATTTAAACGGAAACTTCATATTCAGGCG 2068
295 .....LeuProLeuGlnPheAsnAlaAsnIleThrAlaIleGly 309
2069 GACAAGCGGTGTTCCCGCAATGTTCGCCAAGTGAAGGAGATTGGCAT 2118
309 Lys.....SerValSerPheAspIleHis 316
2119 TTAAGCAATCAGCCCAAGACAGTTTTCGTTGTCGACCGCATCAAGCA 2168
317 ..AlaAsnHisSer.....GlyArgGlyAlaGluLeuLysMe 328
2169 CACATCTGTACAGCTTCGACGTGACGGGTCTGACA.....AGTTGTA 2212
328 tAsnThrIleAsnIleSerAspGlyThrSerLeuThrLeuGlnSerHisV 345
2213 CCGAAAAA.....ACATTAACCGCATTAAGTATGATGCTTCA 2250
345 AlArgLysAspSerAlaPheIleIleSerLysAspLeuThrIleAsnAla 361
2251 TTG.....AGCAAGACCGCAT 2267
362 ThrLysSerAsnPheThrLeuGlnGlnSerProAspSerPheThrAsp 378
2268 CAGAGCAATGTACGCTTCGCCATCAGCGTCATTTAAATCTCAGAGAC 2317
378 sTyrProGlyArgAlaIleSerSerThrLysAsnIleThrIleSerLys 394
2318 TTGCCACACTCAACGCAATCTTAGTCAGCGAGACACGCACTATACG 2367
395 .....GlyAsnValSerLeuGly..... 401
2368 GTTACGGGCAAGCGCACAAACGGCAACCTCAGCTCGTGCGCAATGC 2417
402 .....GlnAsnSerSerSerAspIleLysGlyAsnIle 412
2418 CCAA...GCACATTTAATCAAGCCACATTTAAACGGCAACATCGGCTT 2464
412 eThrIleLysSerSerThrAsnValThrLeuLysAlaIleAsnSerProA 429
2465 CGCAACATGCTTCATTTAATCTAAGC...AACAAAGCGCGTACAAACGGC 2511
429 rGAspPheAlaSerArgThrLeuThrLeuGlnLysLeuAsnValGlyGly 445
2512 AGCTGACGCTTTCCGACAAACGCTAAGCAACGTAAGCAACGTAAGCCATTC 2561
446 AsnLeuThrLeuThrGly.....SerValAlaAspIle 456
2562 CAACGGCAATGTCCTCCATGACCAATTAAGCAATTTCCAT.....T 2602
456 eLysGlyAsnLeuSerIleLeuAsnAspAlaThrPheLysGlyGlyThrS 473
2603 TTGAACACAGCGCTTACCGGAAAAATCAAGCGGCGGCAAGCATACGCGCA 2652
473 eArgIleAsnLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaAsp 489
2653 TTACACTTAAAGACAGCAAGTGAACGCTGCGCGCGC...ACGGAATT 2699
490 IleAsnIleLysGlnGlyValValAsnIleGlnGlyAsnIleThrAsnLys 506
2700 AGCAATTTAAACCTTGACAAACGCGCACATTACATCAATTCGCGCTATC 2749
506 sGlyGlyLeuAsnIle..... 511
2750 GACACGATGCGGAGCGGCGCAACCGCGCAGTGCAGATGCGCGCG 2799
511 ..... 511
2800 CGCGTTCGCGCGCTTCCTATTATCCGTACGCGCCCAACTTCGGCAGA 2849
512 .....ThrThrAsnAlaG 516
2850 ATCCGTTTCACACGCGTGCAGGTAAACGGCAATTAAGCGGACAGGAA 2899
516 naAsnAsnGlnLysThrIle...IleAsnGlyAsnIleThrAsnGlnGly 531
2900 CATTCGCTTATGTGGAACCTTCGCTACCGCAGCGCAATTAAG 2949
532 .....GlyAspLeuAsn 535
2950 CTGCGCGAAAGTCCGAGGCACTTACACCTTGCTGCAACAT..... 2994
536 IleLysAspSerAsnAsnAlaGluIleGlnIleGlyLysAsnIleSe 552
2995 .....ACCGCAACGACCCGTAGTCTCGAGCAATGAGCGTATGAG 3037
552 rGlnLysLysGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrL 569
3038 AA.....GGAAGACAAACACA 3054
569 yslYsIleThrIleLysAlaGlyValAspGluGlyLysSerAspSer 585
3055 CGCGTTCGGAATAATTTCAACCCCTGCAAAACGAACACAGCTGATGC 3104
586 ProAlaSer...AsnAlaAsnLeuThrIleLys..... 595
3105 CGCGCATGGCGTTATACCTTATCCGCAAGACGGCGAGTCCGCTGC 3154
595 ..... 595
3155 ATAATCGGTCAAGACAAAGAGCTTCCGCAAACTGGCAGAGCGGA 3204
596 .....ThrLysThrLeuGlnLeuThrGlyAspLeuAsnIleSerLys 609
3205 GAACAGAGAGCGCGCTTGACGCGCAAAACAGGACCAACTGCGCGCAACA 3254
610 PheAsnLysAlaGluIleThrAlaLysAsnGlyAsnAspLeuThrIleG 626

```

```

3255 ACAGCGGAGAAAGACACGCGCAAGC..... 3282
      ::::: |||||:::
626 YLYSAlaSerAspGluGlnArgGluThrLeuAlaLysLeuGlyVal 643
      ::::: |||||:::
3283 ..CTTGACGCGCTGATGCGCGCC...GGGCGCAATGCGCAC.....GAA 3321
      ::::: |||||::: |||||::: |||||:::
643 allYsAspSerLysIleSerAlaAsnGlyHisAsnValThrLeuAsnSer 659
      ::::: |||||::: |||||::: |||||:::
3322 AAGCGAGAA.....AGTGTGCCGAACCGCGCGCGAGCGCGGCA 3365
      ::::: |||||::: |||||::: |||||:::
660 LysValGluThrSerAsnSerSerSerAlaAspSerSerAsnSpas 676
      ::::: |||||::: |||||::: |||||:::
3366 AAGTCCGCGGATATGCGAGCGGAGAGAG..... 3396
      ::::: |||||::: |||||::: |||||:::
676 nasThrGlyLeuThrIleSerAlaLysAspValThrValAsnAspY 693
      ::::: |||||::: |||||::: |||||:::
3397 .....AAAAACGGGTGCGAGCGGATGAAGACCGGCTTGGCGAAA 3438
      ::::: |||||::: |||||::: |||||:::
693 alThrSerHisLysThrIleAsnIleSerAlaThrThrGlyAsnValThr 709
      ::::: |||||::: |||||::: |||||:::
3439 CAGCGCGACGCGAACC CGCGCGCTACCAACCGCTCCCGCGCGCG 3488
      ::::: |||||::: |||||::: |||||:::
710 ThrLysGluSerThrThrIleAsnAlaIleThrGly.SerValGluValr 726
      ::::: |||||::: |||||::: |||||:::
3489 CCGCGCGCGCGGATTTGCGCGACCGCGCGCGCGCGCGCGCGCGAC 3538
      ::::: |||||::: |||||::: |||||:::
726 hralLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValAsn 742
      ::::: |||||::: |||||::: |||||:::
3539 CGCAGCGCGACCTGATCAGCGGTTATGCCAATAGCGGTTGAGTAATT 3588
      ::::: |||||::: |||||::: |||||:::
743 ValThrAlaThrAspSerLeuThrThrGlnAlaSer..... 754
      ::::: |||||::: |||||::: |||||:::
3589 TCGCCCGACGCTGACAGCGTTTTCGCGTACAGAGCAATGACCGCGT 3638
      ::::: |||||::: |||||::: |||||:::
755 .....SerSerLethrSerSerAsnGlyInThrThrLeuThrAla 769
      ::::: |||||::: |||||::: |||||:::
3639 GTTTCGCGAAGCGCGCGACCGCGCTTGGACAGCGCGATCCGCGGCA 3688
      ::::: |||||::: |||||::: |||||:::
769 ysaSnGlySerIleAlaGlySerIle....AspAlaAlaAsnVglThr 783
      ::::: |||||::: |||||::: |||||:::
3689 CCAAACTACTCCGTTCCGACAGATTCCGCGCTACCGCGCAACAAACGAC 3738
      ::::: |||||::: |||||::: |||||:::
784 LeuAsnThr.....ThrGlyThrLeuThrTh 792
      ::::: |||||::: |||||::: |||||:::
3739 CTGGCGCAATCGGTATCGAGAAAACCTCGCGACGCGCGCG...TCGG 3785
      ::::: |||||::: |||||::: |||||:::
792 rValAlaGlySerAsnIleLysAlaThrSerGlyThrLeuAlaIleAsn 809
      ::::: |||||::: |||||::: |||||:::
3786 CATCCCTGTTTCGACACCGCGCGCGAACCTTCGACGCGCGCATCG 3835
      ::::: |||||::: |||||::: |||||:::
809 lalYsAspAlaLysLeuAspGly.....ThrAlaSer 819
      ::::: |||||::: |||||::: |||||:::
3836 GCAACTCGCGACGCGCTTCCCGCGCTTTCGGGCATACGGCATC 3885
      ::::: |||||::: |||||::: |||||:::
820 GlyAspArg.....ThrValAlaAsnAlaThrAlaSerG 832
      ::::: |||||::: |||||::: |||||:::
3886 GGCAGGTCGACATCGGCATCGCGCGCGCGGTTTATTACGCGCGAG 3935
      ::::: |||||::: |||||::: |||||:::
832 ySerIleSerValThrAlaAlaThrSerSerAsnValAsnIleThrGly 849
      ::::: |||||::: |||||::: |||||:::
3936 CCTTTCAGACG.....GCATCAGAGGCAAAATCC 3964
      ::::: |||||::: |||||::: |||||:::
849 splenSerThrIleAsnGlyLeuAsnIleIleSerLysAsnGlyLysAsn 865
      ::::: |||||::: |||||::: |||||:::
3965 GCGCGCGCGTGTGATACCGCATTCGCAAGATACCGCGCGGTTTC 4014
      ::::: |||||::: |||||::: |||||:::
866 ThrValValLeuLysGlyAlaGluLeuLeuAspValLysTyrIleGlnPro 882
      ::::: |||||::: |||||::: |||||:::
4015 GCGCGATTCGCGATCGCAACCGCATCGCGCGCAACCGCTATTTCGTC 4064
      ::::: |||||::: |||||::: |||||:::
882 yValAlaSerAlaAsnGluValIleGluAlaLysArgAlaLeuGluLysV 899
      ::::: |||||::: |||||::: |||||:::
4065 AAAAGCGGATTAACCGAT.....ACGAAAACGTCAATA 4096
      ::::: |||||::: |||||::: |||||:::

```

```

      ::::: |||||::: |||||::: |||||:::
899 allYsAspLeuSerAspGluGlnArgGluThrLeuAlaLysLeuGlyVal 915
      ::::: |||||::: |||||::: |||||:::
4097 TCGCCACCCCGCGCGCTTCATTCACCGCTACCGCGCGCGCATTAAGCA 4146
      ::::: |||||::: |||||::: |||||:::
916 SerAlaValAlaArgPheValGluProAsnAsnThr..... 926
      ::::: |||||::: |||||::: |||||:::
4147 GATTATTCATTCACACCGCGCGCACATTCATTCATCAGC 4186
      ::::: |||||::: |||||::: |||||:::
927 ....IleThrValAsnThrGlnAsnGluPheThrThrArg 938
      ::::: |||||::: |||||::: |||||:::
seq_name: /SID1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: AAB01826
seq_documentation_block:
ID AAB01826 standard; Protein: 975 AA.
AC
AC AAB01826;
DE 11-SEP-2000 (first entry)
DT
DE Haemophilus influenzae strain J9c HMW2A protein, SHQ ID NO:30.
XX
XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX non-typable Haemophilus influenzae, NTHi; non-encapsulated;
XX recombinant production; Escherichia coli; antibacterial; vaccine;
XX human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.
XX
XX Haemophilus influenzae strain J9c.
XX
XX WO200020609-A2.
XX
XX 13-Apr-2000.
XX
XX 07-Oct-1999; 99WO-CA00938.
XX
XX 07-Oct-1998; 98US-0167568.
XX
XX 08-Dec-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
XX WPI; 2000-303789/26.
XX
XX N-PSDB; AAA52177.
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -
XX
XX Claim 12; Fig 19A-O; 307pp; English.
XX
XX The invention relates to the recombinant production of Haemophilus
XX influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX expression construct used to effect recombinant expression comprises a
XX promoter functional in E. coli (e.g., the T7 promoter) operably linked
XX to a modified hmwABC operon from a non-typable (non-encapsulated) H.
XX influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
XX clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
XX and the hmwB and hmwC genes encode accessory proteins which are
XX responsible for post-translational processing and secretion of the HMWA
XX proteins. The modified hmwABC operon used in the expression construct of
XX the invention contains an A gene modified such that it encodes only the
XX mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
XX and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae
XX strains J9c, KI, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
XX vectors are used for the production of recombinant H. influenzae HMW
XX proteins which can be used as vaccines to mediate a humoral or
XX cell-mediated immune response to provide protection against diseases in
XX humans caused by H. influenzae (e.g., otitis media, epiglottitis,
XX pneumonia and tracheobronchitis). The HMW proteins are also useful as
XX antigens in immunoassays for detecting antibodies against Haemophilus,

```

CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMW protein from a non-typable strain of  
 CC H. influenzae.

XX Sequence 975 AA;

# alignment\_scores:

Quality: 273.00 Length: 1164  
 Ratio: 0.487 Gaps: 54  
 Percent Similarity: 48.110 Percent Identity: 20.619

## alignment\_block:

US-09-303-518D-653 x AAB01826

Align seg 1/1 to: AAB01826 from: 1 to: 975

```

1090 AATGTTCTTATCC.....GAGACAGCAAGAAACC 1121
|||||.....
9 AsnValSerIleGluAsnProSerThrGluArgAsnAspSerAsnGluAs 25
1122 TGTATATATCTGAGGTGGGTCACAGTTATTCAGCCAGACTAATA 1171
|||.....
25 pLeuGluIuThrGlyThrGlyGluAsnIleAsnAsnProGlyValAsna 42
1172 ATGGA.....GAAATATT..... 1185
||
42 snGlnSerIuLysThrLeuThrSerSerIleuGluAsnIleLeuLys 58
1186 .....TCCTTATTTGAC...AAGAAAGATGATGATCTTACAG 1226
|||||.....
59 LysGlySerPheValAsnIleThrAlaThrAspAsnIleGlyValAsnse 75
1227 CAACATGACCAAGG...GGGGCGGTTTGTATTTGAGGTAATTTTA 1273
|||||.....
75 rSerIleAsnIleGlyAspSerGlyHisLeuIleLeuSerGlyGlyGly 92
1274 CGGTCTCGCTTAAACCAAGCAAGGTCGAGCGCGGCTGATATTC 1323
||
92 rG.....AsnGlyGlyValLysIle 99
1324 AGTATGCGACTACCGTTACTTGGAAGTAACGGCTGGCAACGACCG 1373
|||.....
100 AsnLysAsnIleThrSerThr..... 106
1374 CCTGTCCAAATCGCAAGGCGAGCGTGTGTTCAAGCCAAAGG... 1419
|||.....
107 .....GlyGlySerLeuThrIleAsnSerIuGlyTyrPv 118
1420 ..GAAACCAAGGCTCGTCAGCGTGGC..... 1446
|||.....
118 alaSerIleHisSerAsnIleSerLeuGlyThrGlyPheLeuAsnIleThr 134
1447 ..GACGTAAGTCATCTTAGATCAGCAGCGGACGATCAAGGCAAAA 1493
|||.....
135 SerAsnGlySerValAlaPheGlyLysAlaAspLysAspLysAlaArgse 151
1494 ACAAGCCTTTAGTAATCGGCTTGTCAGCGGACGGGAGCGGTGCAC 1543
|||.....
151 rAlaAlaAspAlaGlnIle.....ValAlaGlnGlyIleIleAsnL 165
1544 TGAATGCCGATATCACTTCACCCGACCAAACTATATTGGCTTTGCG 1593
|||||.....
165 euThrGlyGluAsnLys.....ThrPheArg 173
1594 GCGGAGCTTGGATTGAAAGCGGATTCGCTTTCGCCAGCATCA 1643
|||.....
174 LeuAsnAsnValSerIleuAsnGlyValGlyGlnGlyLeuSerIleThrse 190
1644 AATATCCGATGAAGGGCGATGTGTCAACCAACATCAAGACAAAGAA 1693
|||.....

```

```

190 rAsn.....ValGlyAsnGlnThrHisLysPheAspG 201
1694 CCACCGTTACCATTTACAGCAATTAAGATATTACTACACCGGC..... 1737
|||.....
201 LysIleAsnIleThrGlyAsnValThrIleAsnGlnThrAlaProAla 217
1738 .....AATACACACA 1748
|||
218 ThrThrAlaTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSerTh 234
1749 CTGGATAGCAAAAAGAAATTCCTAC.....A 1777
|||.....
234 rLeuAsnValGlnLysAsnSerSerPheThrPheIleLysArgThrGluS 251
1778 ACGGTGCTTTGGCCGAGAAAGATGACCAACAAAGACGGCGGCTCAAT 1827
|||.....
251 eAsnArgPheGlyProThrThrProLeuArgSerSerGlyValPhe 267
1828 CTGAATTACCAACCGGAGAACGGGATCGACCTTCTGCTTCGGCGG 1877
|||
268 PheAsn.....G 270
1878 AACAAATTTAAACGGCAATATC.....ACGCAACAAACGGCAAC 1918
|||.....
270 YThr.....AsnGlyAsnMetValLeuAsnValGlyThrAsnSerArgv 285
1919 TGTTTTTCAGCGGACAGACCGACCGCCTTACATCATTTAGGAAC 1968
|||.....
285 alLeuPheAsnLeuLysProAsnGluAsnThrAsnAsn..... 297
1969 GGGTGTCAAAATGGAAGTATCCACAAAGAGAAATGCTGGGACAA 2018
|||||
298 .....SerLysPro..... 300
2019 CGATTGATCGACCGACCATTTAAACGGAACCTTCATATTCAAGGCG 2068
|||.....
301 .....LeuProLeuGlnPheAsnAlaAsnIleThrAlaIleGlyG 315
2069 GACACGCGGTGTTCCCGCAATGTTCGAAGTGGCAAGCGCATGGCAT 2118
||
315 Lys.....SerValSerPheAspIleHis 322
2119 TTAGCAATCAGCCCGACAGTGTTCGGTGTGCGACCGCATCAAGACA 2168
|||.....
323 ..AlaAsnHisSer.....GlyArgGlyAlaGluLeuLysme 334
2169 CACAATCTGTACACGTTGCGACTGACGGGTGTGACA.....AGTGTGA 2212
|||.....
334 rAsnThrIleAsnIleSerAspGlyThrSerLeuThrLeuGlnSerHisv 351
2213 CGCAAAA.....ACCATTACGAGCATTAAGTGAATGCTGTCA 2250
|||.....
351 alArgLysAspSerAlaPheIleIleSerLysAspLeuThrIleAsnAla 367
2251 TTG.....AGCAAGACCGCAT 2267
|||
368 ThrGlySerAsnPheThrLeuGlnGlnSerProAspSerPheThrAspLys 384
2268 CAGAGCAATGTACAGCTTCGGATCAGCTCATTTAAATCTCACAGGAC 2317
|||.....
384 sTyrProGlyArgAlaIleSerSerThrLysAsnIleThrIleSerGly 400
2318 TTGCCACTCAACGGCAATTTAGTGACGGGAGACACGACATATACG 2367
|||||
401 .....GlyAsnValSerLeuGlyGly..... 407
2368 GTTAGCGCAACGCCACCAAAACGGCAACTCTGCGGCAATGC 2417
|||||
408 .....GlnAsnSerSerAspIleLysGlyAsnI 418
2418 CCAA..GCAACATTTAATCAAGCCATTAACGGCAACACATGGGCTT 2464
|||.....
418 eThrIleLysSerSerThrAsnValThrLeuLysAlaHisAsnSerProA 435

```

```

2665 CGACAAATGCTTCAATTAATCTAAGC...AACAAAGCCGCTACAAAGAGC 2511
      ||| ||||| :|||: :|||: :|||: :|||:
435 rgaspphealaserargthreuthreuglyasnleuasvalglgly 451
      :|||: :|||: :|||: :|||: :|||:
2512 AGTGTGACGCTTTCGACAAAGCCTAAGCAAAAGCTAAGCCATTCGCACT 2561
      :|||: :|||: :|||: :|||: :|||:
452 Asnleuthreuthrgly.....SerValAlaAspI 462
      :|||: :|||: :|||: :|||: :|||:
2562 CAAGGCAATGCTCTCCAGCCGATTAAGGCAATTTCCAT.....T 2602
      :|||: :|||: :|||: :|||: :|||:
462 elysglYasnleuserilleuasnaspalatrhPhelysglYluthrs 479
      :|||: :|||: :|||: :|||: :|||:
2603 TTGAACAGCCGCTTACCGAAATAATCAAGCGGCAAGAGATGCGCA 2652
      ||||| :|||: :|||: :|||: :|||:
479 erglUasnleuasnillethrglyasnphetrhrasnnglyThAlaAsp 495
      :|||: :|||: :|||: :|||: :|||:
2653 TTACACTTAAAGACAGCAATGAGCGCTCGCGGCG...ACGAATT 2699
      :|||: :|||: :|||: :|||: :|||:
496 IleasnileysnglyValValasnleuglnlyasnlethrasnly 512
      :|||: :|||: :|||: :|||: :|||:
2700 AGCAATTAACTTGACAAAGCCACCATTAATTCATTCGCGCTATC 2749
      :|||: :|||: :|||: :|||: :|||:
512 sglYglYleuasnle..... 517
      :|||: :|||: :|||: :|||: :|||:
2750 GACACGATGCGGACGCGCAACCGGACGTGCGGAGATGCGCGCGC 2799
      :|||: :|||: :|||: :|||: :|||:
517 ..... 517
      :|||: :|||: :|||: :|||: :|||:
2800 CGCGCTTCGCGCGCTCCCTATTAATCCGTTAGCGCCCACTTGCGAGA 2849
      :|||: :|||: :|||: :|||: :|||:
518 .....ThrThrAsnAlaGI 522
      :|||: :|||: :|||: :|||: :|||:
2850 ATCCCGTTTCAACGCTGAGGTAAGCGAAATTAAGCGTACAGGAA 2899
      :|||: :|||: :|||: :|||: :|||:
522 nasnasnlnlysthrlle...Ileasnlyasnlethrasnngly 537
      :|||: :|||: :|||: :|||: :|||:
2900 CATTCGCTTATGCGAAMCTCTGCGCTACCGACGCGCAATTAAG 2949
      :|||: :|||: :|||: :|||: :|||:
538 .....GlyAspLeuasn 541
      :|||: :|||: :|||: :|||: :|||:
2950 CTGCGCAAGATTCGACAGCACTTACACCTGCGCTCAACAAAT.... 2994
      :|||: :|||: :|||: :|||: :|||:
542 IleYasnSperasnasnlnalaglulleglnleuglyasnIle 558
      :|||: :|||: :|||: :|||: :|||:
2995 .....ACGGCAACGAACCGTAAGCTCGAGCAATTAAGCGTAGGG 3037
      :|||: :|||: :|||: :|||: :|||:
558 rglnlyslYasnleuthrIleSerSeraspYslIleasnlethrl 575
      :|||: :|||: :|||: :|||: :|||:
3038 AA.....GGAAGACACACA 3054
      :|||: :|||: :|||: :|||: :|||:
575 yslYslleThrlleYasnIaglyAlaSpgluglySeraspSer 591
      :|||: :|||: :|||: :|||: :|||:
3055 CCGCTGTCGAAATCTTAATTTACACCTGCAAAACGACACGTCGATGC 3104
      :|||: :|||: :|||: :|||: :|||:
592 ProAlaSer...AsnAlaAsnleuthrIleYs..... 601
      :|||: :|||: :|||: :|||: :|||:
3105 CGGCGCATGCGGTATACGTTATCCGCAAGAGCGGAGTTCCGCGTC 3154
      :|||: :|||: :|||: :|||: :|||:
601 ..... 601
      :|||: :|||: :|||: :|||: :|||:
3155 ATAATCGGTCAAAGACAGACTTTCCGACAACTCGGCAAGCGGGA 3204
      :|||: :|||: :|||: :|||: :|||:
602 .....ThrlYsthrlleuglnleuthrIglYAspLeuasnIleSercly 615
      :|||: :|||: :|||: :|||: :|||:
3205 GAAACAGAGCGCGCTTACGCGCAAAACAGCACAACTTCGCGCAACA 3254
      :|||: :|||: :|||: :|||: :|||:
616 pheasnlyslalaglullethrlalYasnnglyasnaspLeuthrIleGI 632
      :|||: :|||: :|||: :|||: :|||:
3255 ACAGCGGAAAAAGACAGCGCAAGC..... 3282
      :|||: :|||: :|||: :|||: :|||:
632 yllYslalaseraspglYasnAlaAsnAlaYslYValIthrhleaspYsv 649

```

```

3283 ..CTTGACGCGCTGATTGCGGC...GGGCGCAATGCAAC...GAA 3321
      :|||: :|||: :|||: :|||: :|||:
649 allYasnSperlyslIleSerAlaasnnglyHlsAsnValIthrlleuasnr 665
      :|||: :|||: :|||: :|||: :|||:
3322 AAGGCGAA.....AGTGTCCGACACCGCGCGGCAAGCGGCGGGA 3365
      :|||: :|||: :|||: :|||: :|||:
666 lysValgluthrSerasnSeraspSerAlaAspSeraspSeraspas 682
      :|||: :|||: :|||: :|||: :|||:
3366 AATGCGCGCATTAATGACGCGGACGAGAG..... 3396
      :|||: :|||: :|||: :|||: :|||:
682 nasnthrglyleuthrIleSerAlaYasnValIthrlalasnaspY 699
      :|||: :|||: :|||: :|||: :|||:
3397 .....AAAAACGGGTGACAGCGGATTAAGACACCGCTTGCGCAAA 3438
      :|||: :|||: :|||: :|||: :|||:
699 alIthrsrlslYsthrlleasnIleSerAlaThrlYslYasnValIthr 715
      :|||: :|||: :|||: :|||: :|||:
3439 CAGCGGAGCGGAAACCGCGCGCTACACCGCGCTTCCCGCGCGCG 3488
      :|||: :|||: :|||: :|||: :|||:
716 ThrlysgluserthrlleasnAlaIatrhrglySerValgluValT 732
      :|||: :|||: :|||: :|||: :|||:
3489 CGCGCGCGCGGATTTGCCGCAACCGGACCGCAACCGCAACCCCAAC 3538
      :|||: :|||: :|||: :|||: :|||:
732 hrAlaYsthrglyaspIleSerclyThrlleSerclyasnThrlalasn 748
      :|||: :|||: :|||: :|||: :|||:
3539 CGCAGCGCACTGATCAGCCGTTATGCAATAGCGGTTGATGAATTT 3588
      :|||: :|||: :|||: :|||: :|||:
749 ValIthrlatrhAspserleuthrIthrglnAlaser..... 760
      :|||: :|||: :|||: :|||: :|||:
3589 TCCGCAACGCTCAACAGCGTTTCCGCGTACAGGCAATTCGACCGGT 3638
      :|||: :|||: :|||: :|||: :|||:
761 .....SerSerIleThrsSerasnnglylnthrlleuthrAlaL 775
      :|||: :|||: :|||: :|||: :|||:
3639 GTTGGCGAAAGACCGCGCACGCGCTTGGACAAAGCGGATTCGGGACA 3688
      :|||: :|||: :|||: :|||: :|||:
775 yasnnglySerIleAlaYserIle.....AspAlaIasnValIthr 789
      :|||: :|||: :|||: :|||: :|||:
3689 CCAACACTACCGCTTGCAAGATTTCCGCGCTACCGCGCAACAAACGAC 3738
      :|||: :|||: :|||: :|||: :|||:
790 leuasnthr.....ThrglyThrlleuthrTh 798
      :|||: :|||: :|||: :|||: :|||:
3739 CTGCGCAAAATCGGTATGCAAAAAACCTGGAGCGCGCGG...TCGG 3785
      :|||: :|||: :|||: :|||: :|||:
798 rValAlaYserasnIleYasnIatrhserglYthrlleuAlaIleasnA 815
      :|||: :|||: :|||: :|||: :|||:
815 lalYasnspIatYslleuaspgly.....Thrlaser 825
      :|||: :|||: :|||: :|||: :|||:
3836 GCAACTCGGACGCGCTTGCCACAGGTGCGGTTTCGGGCAATACGGCATC 3885
      :|||: :|||: :|||: :|||: :|||:
826 GlyaspArg.....ThrlValasnAlatrhAsnAlaSerGI 838
      :|||: :|||: :|||: :|||: :|||:
3886 GGCAGGTGACATCGCATACGACGCGGCGGCTTATAGCGGCGAG 3935
      :|||: :|||: :|||: :|||: :|||:
838 yserglYserValIthrlalalathrSerSerasnValasnIleThrglyA 855
      :|||: :|||: :|||: :|||: :|||:
3936 CCTTGACAGC.....GCATCAAGCGCAAAATCC 3964
      :|||: :|||: :|||: :|||: :|||:
855 spleuserthrlleasnnglyleuasnIleIleSerclyasnnglyYasn 871
      :|||: :|||: :|||: :|||: :|||:
3965 GCGCGCGCGCTGCTCATTAAGCATTCAGGCAAGATACCGCGAGGTTTC 4014
      :|||: :|||: :|||: :|||: :|||:
872 ThrValValleuYslYalaglulIleaspValIlystrYIleIntrGI 888
      :|||: :|||: :|||: :|||: :|||:
4015 GCGGATTCGCAATCGACCGCATCGGCGCAACCGGCTATTCGTCA 4064
      :|||: :|||: :|||: :|||: :|||:
888 yValAlaserAlaasngluValIleglunlalYsatglAlaleuGlulYsv 905
      :|||: :|||: :|||: :|||: :|||:
4065 AAAAGCGGATTAACGAT.....ACGAAACGTCATAA 4096
      :|||: :|||: :|||: :|||: :|||:
905 allYasnSperSeraspglugluargluthrleuAlaYslleuglyVal 921
      :|||: :|||: :|||: :|||: :|||:
4097 TCGCCACCGCGCGCTTGCAATTCACCGGCTACCGCGGCGCATTAAGGA 4146

```



259 LysAsnGlnAspAspIleThrAsnArgLysLeuAsnGlyIleThr..... 273  
958 GTATTTCACGACCATCAATAATGGAAATFACCTTTTACCACAAATA 1007  
274 ..... 111111 : : : :  
1008 TAATGGCGCAGAAATAATGCATGCCAAACATAAACACTATTCTCTACCTT 1057  
278 nTth..... 279  
1058 ATAGATTAAAAACAGAACCGTTCAATTGTTTAATGTTCTTATCCGAG 1107  
280 ..... PheAsnIleAlaGlnGlySer 286  
1108 ACACGACGAGACCTGTTTATCATGCTGCAGGTGGGTCAACAGTTTCG 1157  
287 ThrAlaAsn..... PheHisIleLysThrSerValMetThrProly 300  
1158 ACCGAGACTGAAT.....AATGGAAATAATTTCTTTATTTG 1195  
300 sProAsnSerAsnTyralAlaLeuPheAsnGly..AsnIleSerValLeuG 316  
1196 ACAAGA.....AAGGTGAATTGATTACTTACCACCAACATCAACCA 1239  
316 LysGlyGlyThrValAsnPheGluLeuAsnAlaSerSerThrHisThr 332  
1240 GCGCGCGGCGGTTGTATTGAGGTAATTTTACGGTCTCGCCTAAAA 1289  
333 ThrSerGlyAlaIleIleAsnSerGlnAspHe..... 343  
1290 CAAGGAAACGTGGCAAGCGCGGCGTTCAATATCATGATGACGATACG 1339  
344 .....AsnValSerGlyLysLysL 351  
1340 TTACTTGAAGTAAACGCGCTGGCAACACGCCCTGTCCAAATAGCGC 1389  
351 euAsnLeuLysAlaSerGlySerThrAsnThrAlaPhe.....LeuIle 365  
1390 AAGGACGCTGCTGTTCAAGCCAAAGGGGAAAC..... 1425  
366 LysAsnAsnLeuThrLeuAsnAlaThrGlyLysAsnIleGluIleLysG 382  
1426 ....CAAGGCTCGCTACGCTGGCGGCGGTAAGTCATCTTACATCAGC 1471  
382 nValGlnGlyThrAspSerArgIleGlnLysGlyValValAlaGlnGln 399  
1472 AGCGGACGATCAAGGCAAAAAACAAGCCTTATGTAATGCGGTGTC 1521  
399 snIleIlePheGluGly.....GlyAsnIleThrLeuGly 410  
1522 AGCGGACGAGGGGACGTGCAACTGAAATCGATATATCATGTAACCCGA 1571  
411 SerGlnLysAlaProThrGluIleLysGlyAspValThrValLys.... 425  
1572 CAACCTATTTCGCTTCCGCGGCGGACGTTTGGAATTTGACGGGCATT 1621  
426 .....GlnGlyThrAsnAlaThrLeuArgSerAla 436  
1622 CGCTTTCGTCACCGCATTCAAATATACGATGAAGGGGCGATGTC 1671  
436 snPheAspAsnHisLys.....GlyAlaLeuIleVal 446  
1672 AACCAATCAAGACAAAGAAATCCACGCTTACATTCAGGCAATTAAGA 1721  
447 AsnGlyAsn.....ValThrAlaAsnGlyAs 455  
1722 TATTCTCAACACCGCAAT.....AACACAACTTGGATGCA 1759  
455 nLeuThrAlaAspGlyAspThrIleLysIleLysGlyAsnLeuAspVal 472  
1760 AAAAAGAAATTCCTCAACAGGTGGTGGGCAAGATGCAACCA 1809  
472 IagIlnGlyAlaLysPheAsnGly.....SerThrLys 482  
  
1810 ACCGACGGCGGCTCATCTGAATTACCAACCGGAGAGCGATGCGAC 1859  
483 AsnAsnLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrSerIleI 499  
1860 TTTACTGCTTTCGCGGAGCAAAATTTAAACGGCAATATCAACGCAACAA 1909  
499 eaSplIleThrGlnGlyValValAsnLeu...GlyAsnValThr...Asn 514  
1910 ACGGCAACTGTTTTCACGCGGACGACACCGCCGCTACCAATCAT 1959  
514 spGlyLysLeu.....AsnIleThrThrHisAlaLys..... 524  
1960 TTAGGAAGCGGCGGTCAAAATGGAAGATATCCCAAGAGAAATCGT 2009  
525 .....SerGlyGlnLysSerIleIleArgIlyAspIleI 536  
2010 GTGGGCAACGATTTGGATGACCGGCACATTTAAAGCGGAAATTCAT 2059  
536 eaSnLysGlnGlyAsnLeuAsnIleThrAspAsnAsnSerAlaGlnI 553  
2060 TTCAGGCGGACAGCGGTGTTCCCGCAATGTTGCCAAAGTGAAGCG 2109  
553 LeuL.....IleGlyGlyAsnIleSerGlnLysGluGly 564  
2110 GATTGCAATTGAACATCAACGCCCAAGCAGT..... 2142  
565 AsnLeuThrIleSerSerAspLysValAsnIleThrLysGlnIleThrI 581  
2143 ....TTGGTGTGCGACCGCATCAACGACACATCTGTACACGTTGG 2188  
581 euLysAlaGlnVal...AspGlyLysSerSerSerSerThrAlaSer 597  
2189 ACTGACGGGTGTCACAGTTGTACCGAAACCAATTCACGAGATAA 2238  
597 spAlaAsnLeuThrIleLysThrLysGluLeuThrPheThrAspLysn 613  
2239 GTGATTGCTTCATTTAGCAAGACCGACATCAAGGCAATGACGCTTGC 2288  
614 AsnIleSerGlyPheAsnLysAlaGlnIle.....ThrAlaLys 626  
2289 CGATCACGCTCATTTAAATCTCACAGAGCTTGCACACACGCAACGCAAT 2337  
626 saPaAsnSerAspLeuIleIleLysAlaSerSerAspAsnSerAsn 643  
2338 .....CTTAGTGACGCG 2349  
643 IalysGlnValThrPheAspLysValLysAspSerLysIleSerAlaGly 659  
2350 GGAACACGCACTATACGTTACGCGCAACGCC.....ACCCAAA 2390  
660 .....AsnHisAsnValThrLeuAsnSerLysValGluThrSerAs 673  
2391 C.....GGCAACCTGACCTCGTGGCAATGCCAACGA 2425  
673 nSerAspGlySerThrGlyAsnLysSerAspAsnAsnIleGlyLeuT 690  
2426 CATTTAATCAAGCCACATTAACGCGACACATCGGCTTGACATGCT 2475  
690 hIleSerAlaLysAspValThrValAsnSerAsnIleThrSerHisLys 706  
2476 TCATTTAATCTAAGCAACACGCGCTACAAACGACAGCTGACGTTTC 2525  
707 ThrValAsnIleSer.....AlaSerGlnGlyIleThr..... 718  
2526 CGACACGCTAAGCAAGGTAAGCATTCGACATCAACGCAATGCT 2575  
719 .....ThrLysAlaGlyThrThrIleAsnAlaThrThrGlySerValG 733  
2576 CCTAGCGATTAAGCAGATTCATTTGAAAACAGCGCTTACCGGA 2625  
733 IuValThrAlaLys.....ThrGly 739



```

2626 AAATCAGCGC.....GCCAGGATACGGCATTACCTTAAGAA 2666
      |||||
740 AsptleSerGlyThrIleSerGlyValSerValThrAlaSerTh 756
2667 CAGCAATGAGCGCGCGCGCGG..... 2691
756 rGlyAspLeuThrValArgLysAlaAlaThrIleSerValThrGluGly 773
2692 .....ACGGATTAGCAATTTAAACCTTGACACGCCACCAT 2730
      |||||
773 IaAlaThrLeuThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySer 789
2731 ACACCTCAATTCGCCCTATTCACACAGCGCGGCGCAACCGGAG 2780
      |||||
790 SerIleThrSerThrLysGlyGlnValAspLeuSerIleGlnAspGlySe 806
2781 TGCGGCAGAT.....GGCGCGCGCGCGCGCGCGCGCGTCC 2818
      |||||
806 rIleAlaGlyGlnIleSerAlaAlaAsnValThrLeuAsnThrThrGly 823
2819 TATATCCGTTACGCCCGCACTTCGAGAGATCCGTTCAACAGCGTG 2868
      |||||
823 hrLeuThrThrValGluGlySerAsnIleLysAlaThrSerGlyThrLeu 839
2869 ACGGTAACGGCAATTTGACGCGTACGGAGACATTCGCTTATGTGGA 2918
      |||||
840 AlaIleAsnAlaLys..... 844
2919 ACTCTTCGGGTACCGGCGCAATTTGAACCTGGCGAAAGTCCGAG 2968
      |||||
845 .....AspAlaLysLeuAspGlyThrAlaSerIleLysAsn 856
2969 GCACCTACACCTTGCTGTCAACATACCGGCAACACCGCTAAGCTC 3018
      |||||
856 rGthrGluValAsnAlaThrAsnAlaSerGlySerValThrAla 872
3019 GAGCAATTTGACGTTAGTGAAGAAACACACACCGCTGTCCGAAA 3068
      |||||
873 LysThrSerSerAsnValAsnIleThrGlyAspLeuSerThrIleAsnG 889
3069 TCTTAATTTACCTCGCAAAACGAAACGTCGATCGCGCGCGATGGCT 3118
      |||||
889 rLeuAsnIleIleSerGluAsn.....GlyArgAsnThrV 901
3119 ATCAGCTTATCCGCAACGAGCGGAGTTCCGCTGATATTCG..... 3162
      |||||
901 alaArgLeuArgGlyLysGluIleAspValLysTyrIleGlnProGlyVal 917
3163 .....GTCAAGACACA 3173
918 AlaSerValGluGluValIleGluValAlaLysArgValLeuGluVal 934
3174 AGACCTTCGACAAA.....CTCGGCAAGCGCGGGAACACAG 3211
      |||||
934 sAspLeuSerAspIleGluArgLysThrLeuAlaLysLeuGlyValSer 951
3212 AGCGCGCTTGAAGGCAAAACACACACTTCGCGGCAACACAGCGG 3261
      |||||
951 lAlaValArgPheValGluProAsnAsnAlaIleThrIleAsnThrGln 967
3262 GAAAAAGACACAGCGCAACGCTTGACCGCGCTGCGCGCGCGCGC 3309
      |||||
968 GluPheThrThrArgProSerSerGlnValIleIleSerGluGlyLys 983
seq_name: /sidsi/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AA01840
seq_documentation_block:
ID   AAB01840 standard; Protein; 1010 AA.
XX
AC   AAB01840;
XX
DT   11-SEP-2000 (first entry)
XX

```

```

DE   Haemophilus influenzae strain PMH1 HmW2A protein, SEQ ID NO:55.
XX
XX   HmW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
XX   non-typable Haemophilus influenzae; NTHi; non-encapsulated;
XX   recombinant production; Escherichia coli; antibacterial; vaccine;
XX   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX   detection; diagnosis.
XX
OS   Haemophilus influenzae strain PMH1.
XX
PN   WO20020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999; 99WO-CA00938.
XX
PR   07-OCT-1998; 98US-0167568.
XX   08-DEC-1998; 98US-0206942.
XX
PA   (CONN-) CONNUGHT LAB LTD.
XX
PI   Loosmore SM, Yang Y, Klein MH;
XX
DR   WPI, 2000-303789/26.
XX   N-PSDB; AAA52189.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
XX   proteins of Haemophilus which are used as a vaccine to provide
XX   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 25A-O; 307pp; English.
XX

```

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the CC HMW proteins can be used to isolate and clone hmw genes from other CC non-typable strains of Haemophilus via hybridisation reactions. The present sequence represents an HMWA protein from a non-typable strain of H. influenzae.

Sequence 1010 AA:

alignment\_scores:

Quality: 272.00 Length: 1116  
Ratio: 0.514 Gaps: 49  
Percent Similarity: 47.401 Percent Identity: 20.878

alignment\_block:

US-09-303-518D-653 x AAB01840 ..

Align seg 1/1 to: AAB01840 from: 1 to: 1010

307 GCATTGGCGCGCATCAATATATTGTGAGCGTGGCACATACGCGGCTA 356







```

707 ngIuLeuSerLeuSerThrArgGlySerLeuLysAsnSerHisThrLeuG 724
977 AAATGGGAATACTCTTTTAAACGACATATATAT..... 1011
724 InAlaGlyLysArgIleArgIleLysAlaAsnLeuAspAsnAlaIa 740
1012 .....GGCGAGAAAAATCGATGCGCAAAACATAAACCTA 1046
741 GInGlyAsnIleGInSerGlyGlyThrThrAspIleGlyThrGInHis.. 756
1047 TTCCTACCTTATAGATTAAACACGAAACCGTTCAATGTTTAAATGTTT 1096
757 .....AsnLeuThrAsnArg 762
1097 CTTTATCCGACACAGACAGACACTGTTTATCATGCTGCGAGGGGGTGC 1146
762 LysLeuIleAspGlyGInGInThrLysIle.....GInAlaGlyGInMet 776
1147 AACAGTTATCGACCCAGACTGATATATGAGAAATATTTCTTTATTT.. 1194
777 AsnAsnIleGlyThrGlyArgIleArgIleTyrGlyAspAsnIleAlaIleAla 793
1195 .....GACAAAGCAAAAGTGGAATGTACTTAA 1222
793 aThrArgLeuAspAsnGInAspGInAsnGlyThrGlyAlaAlaIleAla 810
1223 CC...AGCAACATCAACCAAGCGCGGGCTTTGTATTTTGAAGGCTAAT 1269
810 LAspGInAsnLeuAsnLeuGlyIleGlyGInLeuAsnAsnArgGInAsn 826
1270 TTTACGGTCTCGCTTAAACAAACAGAAACGTGGCAAGCGCGGGCTTCA 1319
827 SerLeuLeuTyrSerGlyAsnAspMetAlaValGlyAla..... 840
1320 TATCAGTATGGGAGTACCGTTACTTGGAAAGTAAAGCGGTGGCAACG 1369
841 .....LeuAspThrAsnGlyGInAlaIleArg 849
1370 ACCGCTGTCCAAATCGGCAAGGCGACGCTGCTGTTCAAGCCAAAGG 1419
849 LysLysAlaGInArgIleHisAsnAlaGlyAlaThrIleGlyAlaIleAlaGly 865
1420 GAAACACAA...GGCTGGTCAGCGTGGCGGAGCGGTAAAGTCATCTTGA 1466
866 LysSerThrLeuGlyValGlyLysLeuHisAsnThrAsnGInHisLeuGly 882
1467 TCAACAGCGCGAGCATCAAGCAAAAAA.....CAAGCCT 1501
882 sThrGInLeuValGInThrGlyArgGInHisIleValAspTyrGInAla 899
1502 TTAGTGAATCGGCTTGGTCAGCGGACGGGAGCGGTGCAACTGAATGCC 1551
899 heGlyArgHisGlyLeuLeuArg...GlyGlyThrGInHisGlyLeuGly 914
1552 GATATCAGTTCAAACCCGACAACTCTATTTGGCTTCGGCGCGGAGC 1601
915 TrpSerValTyrAsnAspGInSerAspHisLeuArgThrProAspGlyAla 931
1602 TTTGGATTGAAACGGCATTCGTTCTGTTCCACCGCATT...CAAAATA 1648
931 aAlaHisGInAsnThrHisLysTyrAspTyrGlyLysValThrGInLys 948
1649 CGCAT.....GAAGGGCGCATGTTGTCAACACACATCAAGAC 1686
948 hrGInValThrGInThrAlaProAlaLysIleLeuSerGlyAsnAsp... 963
1687 AAAGAAATCCACGTTACCATTAACGCAATTAAGATATTACTACA..... 1731
964 .....LeuThrIleAspGlyLysGlyValAlaPheAsnThrAsp 976
1732 .....ACCGCAATTAACAAACCTTGATTACCAAAAAAAGAAATTG 1771

```

---

```

976 rGInIleIleAlaGlyLysAsnLeuIleValGInThrGlyLysAspGlyL 993
1772 CTAACAAC...GGTGGTTTGGCGAGAAAGATCAACCAAAACGACGGG 1818
993 euHisAsnGInGInThrPheGlyGlyLysLysValPheSerGlyAsnGly 1009
1819 CGGCTCAAT..... 1827
1010 LysLeuHisSerTyrTrpArgGlyLysLysGlyArgAspSerThrG 1026
1828 .....CTGAATTACCA...CCGGAAAGAGCGGATCGCACTTAC 1864
1026 yHisSerGInGInAsnTyrThrLeuProGInGInIleThrArgAsnIle 1043
1865 TGCTT..... 1869
1043 erLeuGlySerPheAlaTyrGlySerHisArgLysAlaLeuSerHis 1059
1870 .....TCCGGCGGACAAATTTAAACGGCATATCAGCAACAACG 1913
1060 AlaProSerGInGlyThrGInLeu.....ProGInSerAsnG 1072
1914 C.....AACTGTTTTCAGCGGACGACAGCCACACCGC 1945
1072 yLysSerLeuProTyrThrSerAsnSerPheThrProLeuProSerS 1089
1946 ACCCGTAC...AATCATTTACAGCGGGTGTCAAAATGGAAGT 1989
1089 erLeuTyrIleIleAsnProValAsnLysGlyTyr...LeuValGInThr 1104
1990 ATCCACAA...GGAGAAATCGTGTG...GACAAAGATTGATGCA 2030
1105 AspProArgPheAlaAsnTyrArgGInThrLeuGlySerAspTyrMetLe 1121
2031 CGGCATTTTAAAGCGGAA...AACTTCATATTACAGCGCGCA... 2070
1121 uAspSerLeuLysLeuAspProAsnAsnLeuHisLysArgLeuGlyAspG 1138
2071 .....CAAGCGTGTTCGCCGAATGTTGCCAAAGTGGAAAGC 2109
1138 LYTyrTyrGInGInArgLeuIleAsnGInGInIleAlaGlyLeuThrGly 1154
2110 GATTGCGATTTA..... 2121
1155 HisArgArgLeuAspGlyTyrGInAsnAspGInGInPheLysAlaIle 1171
2122 ...AGCAATCAAGCCCAAGCAGTT.....TTGGGTG 2149
1171 uMetAspAsnGlyAlaThrAlaAlaArgSerMetAsnLeuSerValGlyI 1188
2150 TCGCACCGCATCAAGCCACACAAATCTGTACCGTTGCGAGTGGACGGGT 2199
1188 LeAlaLeuSerAlaGInGInValAlaGInLeuThrSerAspIleValItr 1204
2200 CTGACAAGTTGTACGAAAAA.....ACCATTAC 2228
1205 LeuValGInLysGlyValLysLeuProAspGlyGlyThrGInThrValLe 1221
2229 CGACGATAAGTGATGCTTCATTAGCAAGACGCACTCAGAGCAATG 2278
1221 uValProGInValTyrValArgValLysAsnGlyAspIleAspGlyLysG 1238
2279 TCAGCCTT.....GCCGATCAGGCTCAATTAAATCTCAGAGCACTTGC 2322
1238 LysAlaLeuLeuSerGlySerAsnThrGInIleAsnValSerGlySerLeu 1254
2323 ACATCAACGGCAATCTTAGTCAGCGGAGACACGCACTATACGGTTAC 2372
1255 LysAsnSerGlyThrIleAlaGly..... 1262
2373 GCGCAACGCC.....ACCAAAACGGCACTCAGGCTGTGGC... 2412
1263 ArgAsnAlaLeuIleIleAsnThrAspThrLeuAspAsnIleGlyGlyAla 1279

```



```

1792 snAlaGlyAlaAlaVal...serPheGlyGlnGlyLysTrpSerLeuGly 1807
3904 ATACGGCGGGCGGGCGGTTTAGTAGCGGCGACCTTTCAGCGGCATCAG 3933
1808 VAlTrAlaGlyGlyAsnValGlyLysGlyTyr.....GlyAsnG1 1821
3954 AGGCAAAATCCCGCGCGTGTGCATTTAC.....GGCATTCAG 3994
1821 yAsPserIleThrHisArgHisSerHisIleGlyAspLysGlySerGlnT 1838
3995 CAAGTACCGCGCAGGT.....TTCCGCC 4017
1838 hrLeuIleGlnSerGlyAspThrThrIleLysGlyAlaGlnValArg 1854
4018 GGATTCCGCGATCGACATCGGCGGCGCATTTTCGTCGA.. 4065
1855 GlyLysGlyValGln.....ValAsnAlaLysAsnLeuSerIleGlnSe 1869
4066 .....AAACGGATTACCGCATAGAAACGTCATATCGCC... 4101
1869 rValGlnAspArgGlnThrTyrGlnSerLysGlnGlnAsnAlaSerAlaG 1886
4102 .....ACCCGCGCGCTTCATTCACCGCTACCGCGCGGCATTAAGCA 4146
1886 lValThrValGlyTyrGlyPheSer.....AlaGly.....Gly 1897
4147 GATTATTCATTCAAACCG.....GGCGACACATTCCTCATACCGCCTTA 4190
1898 AspIyrSerGlnSerLysIleArgAlaAspHisValSerValThrGluG1 1914
4191 TTTCAGC..... 4197
1914 nSerGlyIleTyrAlaGlyGlnAspGlyTyrGlnIleLysValGlyAsnH 1931
4198 .....CTGCCATATCCGATGCCGCTTCGCCGC 4224
1931 lstrAspLeuLysGlyGlyIleLeuThrSerThrGlnSerAlaGlnAsp 1947
4225 AAATCCGAACGCGCGTCATACCGCGTATGGCGGCGAGATTCGGCA 4274
1948 LysGlyLysAsnArgPheGlnThrAlaThrLeuThrHisSerAspIleLy 1964
4275 AACCCGCGATGCG.....GAATGGGCGGTAAACCGCG 4306
1964 sAsnHisSerGlnTyrLysGlyLysSerPheGlyLeuGlyAlaSerAlas 1981
4307 AAATCAAAAGTTTCAGCGTGCCTCCACCGTCCCGCCCAAGGCGCG 4356
1981 erIleSerGlyLysThrLeuGly.....GlnGlyAlaGlnAsnLysPro 1995
4357 CAA.....TTGAAAGCGCAGCAGCAGCGCGGCATCA 4388
1996 GlnAsnLysHisLeuThrSerValAlaAspLysAsnSerAlaSerSe 2012
4389 ATTAGCGTAC 4398
2012 rValGlyTyr 2015
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.AAB01837
seq_documentation_block:
ID AAB01837 standard; Protein: 1073 AA.
AC AAB01837;
XX
XX 11-SEP-2000 (first entry)
DE Haemophilus influenzae strain LCD02 mature HMW2A protein, SEQ ID NO:49.
XX
XX Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTM1; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;

```

```

KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
XX Haemophilus influenzae strain LCD02.
XX
XX WO200020609-A2.
XX
XX 13-APR-2000.
XX
XX 07-OCT-1999; 99WO-CA00938.
XX
XX 07-OCT-1998; 98US-0167568.
XX
XX 08-DEC-1998; 98US-0206942.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Loosmore SM, Yang Y, Klein MH;
XX
XX WPI: 2000-303789/26.
XX
XX N-PSDB; AAA52186.
XX
XX
XX Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -
XX
XX Claim 8; Fig 23A-P; 307pp; English.
XX
XX
XX The invention relates to the recombinant production of Haemophilus
XX influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX expression construct used to effect recombinant expression comprises a
XX promoter functional in E. coli (e.g., the T7 promoter) operably linked
XX to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
XX influenzae (NTM1). Most HMW-expressing NTM1 strains contain two hmw gene
XX clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
XX and the hmwB and hmwC genes encode accessory proteins which are
XX responsible for post-translational processing and secretion of the
XX proteins. The modified hmwABC operon used in the expression construct of
XX the invention contains an A gene modified such that it encodes only the
XX mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
XX and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
XX strains Joyce, K1, K21, LCD02, PMH1, 15 and 12. The nucleic acids and
XX vectors are used for the production of recombinant H. influenzae HMW
XX proteins which can be used as vaccines to mediate a humoral or
XX cell-mediated immune response to provide protection against diseases in
XX humans caused by H. influenzae (e.g., otitis media, epiglottitis,
XX pneumonia and tracheobronchitis). The HMW proteins are also useful as
XX antigens in immunoassays for detecting antibodies against Haemophilus,
XX HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
XX HMW proteins can be used to isolate and clone hmw genes from other
XX non-typeable strains of Haemophilus via hybridisation reactions. The
XX present sequence represents a mature HMW protein from a non-typeable
XX strain of H. influenzae.
XX
XX Sequence 1073 AA;

```

```

alignment_scores:
Quality: 269.50      Length: 1095
Ratio: 0.496        Gaps: 54
Percent Similarity: 49.589   Percent Identity: 20.639

```

```

alignment_block:
US-09-303-518d-653 x AAB01837 ..

```

```

Align seg 1/1 to: AAB01837 from: 1 to: 1073

```

```

787 TCATTGGCGACAGTGGCTACCAATGTTTCTATGATGCCCAAGCA 836
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
72 AsnIleGlyAspSerGlyHis...LeuThrLeuTyrLysLysArgLysAs 87
837 AAAGTGTTAAATTAATGGCGTATTGCAAAACAGCACCCCTATATAGAA 886
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

```

87 nary.....SeraspGlyIleGlnIleasnlySaspIleThrSert 101  
 887 AAGCAATGGCTTCAGCTAGTTCGTAAGATTGG.....TTCATGAT 930  
 101 hcglygIySerLeuThrIleasnSerAspSptPvalAspIleHisgIy 117  
 931 GAATCTTTCGTCGATACCATTCATGATTCCTACGACCAACATCAAA 980  
 118 AsnIleThrLeuGlyGluGly.....PheLeuasnIleThrSertSe 131  
 981 TGGGAATACTTTTAAACGACATATAATGCGCGAGAAAAATCGATG 1030  
 131 raspSerValAlaPheGluGlyGlyasnGlyasnlySgIyArgSerSera 148  
 1031 CCAACATMAA.....CACTATTCCTACCTTATAGATTAAAAACA 1071  
 148 laseRlaGlnIleIlealagInGlyThrIleThrLeuThrGlyLuasn 164  
 1072 CGAACGGTTCATTTGTTTAACTGTTCTTATCCGAGACAGCAAGAACCC 1121  
 165 LysThrPheArgLeuasnSasnValSerLeuasnGlyThr..... 177  
 1122 TGTTATCATGCTGCAGTGGGGTCAACAGTTATCGACCCAGCTGATA 1171  
 177 ..... 177  
 1172 ATGGACAAATATTCTCTTATGACAAAGAAAGTGATGATGACTT 1221  
 178 ..GlyasnGlyLeuSerIleIleSer.....Thr 186  
 1222 ACCGACAACTCAACCAAGCGCGCGGT...TGTATTTTGGAGGTAA 1268  
 187 AlSerAsnLeuSerHisArgLeuaspGlyGluIleasnValSerGlyasn 203  
 1269 TTTTACGGTC.....TCGCCTAAAAACAACGAACCTGGCAAGCC 1308  
 203 nValThrIleasnGlnThrThrGlnGlnAsnIleGluTyrrIlySAlas 220  
 1308 ..... 1308  
 220 eRserAspSertTyrrTPasnValThrSerPheasnLeuArgIuaspSer 236  
 1308 ..... 1308  
 237 LysPheThrPheIleLysTyrrValAsnSerAlaArgasnGlyaspValar 253  
 1309 .....GCGGGCGTT..... 1317  
 253 gGlyArgSerPheAlaGlyValIlePheasnAlaLysGlyLeuThrThrs 270  
 1318 .....CATATCAGTATGGCAGTACCTTCTGGAAGTA.....AAC 1356  
 270 erPheasnValLysLysGlySerThrValasPheLysLeuLysProasn 286  
 1357 GCGGTGCAACGACGCC.....CT 1376  
 287 SerGlyArgSerGlnLysArgIleProIleGlnPheGlnSerAsnIle 303  
 1377 CTCCAAAATCGGCAAGCAGCGCTGCTGTCACCAAGGGGAAAAACC 1426  
 303 eSerValSerGlyGlyArgValAsnIleasnThrLeuAlaasnLeuThr 320  
 1427 AAGGC.....TCGTCACGCTGGGCGAGCGT 1452  
 320 hrgIyGlyGlyValGluIleArgSerSerSerIleasnValSerAspGly 336  
 1453 AAAGTCATC...TTGATCAGCAGCGGAGAGATCAAGCAAAAAACAAGC 1499  
 337 SerThrLeuSerMetThrAlaGlnAlaArgspArgAsnAlaPheGluIle 353  
 1500 CTTTAGTCAATCGGCTTGGTCACGCGCAGGGGAGCGTGCAACTGATG 1549  
 353 eThrLysAspLeuValIleasnAlaSerAsnSerAsnLeuSerIleIleG 370

1550 CCGATATACGTTCAACCCGACAAACTATTTGCGCTTTTCGGGCGGA 1599  
 370 IngIlnAsnAsp.....GlyPhe..... 375  
 1600 CGTTTGGATTGTAACGGGCAATTCGCTTTCGTCACCGCATCAAAAT 1647  
 376 .....AspAsnAsnGlnLysAlaAsnAlaIleasnSerLysTyrrAsnVa 390  
 1648 .ACCGATGAAGGGCGATGATTCACACCAATCAACAGCAAAAGATCA 1696  
 390 ThrIleGlnGlyLysnValThrLeuGlyGlyGlnAsnSerSertSert 407  
 1697 CCGTTAC.....ATTACAGCAATTAAGATTAATCTACACAC 1734  
 407 ThrIleThrGlySerValasnIleGlyAlaAsnAlaAsnValThrLeuGln 423  
 1735 GGCATATAACAACACTTGATATGCAAAAAGAAATTCCTTCAACGGTTG 1784  
 424 AlaHisasnGlyAsn...AspArgasnLysLysLeuThrPheGlyAsn.. 438  
 1785 GTTTCGCGAAGAAGTGCACCAAAAGCAAGCGGCGCTCAATCTGAAAT 1834  
 439 .....ValSerValGluGlyGluLeuArg..... 446  
 1835 ACCAACCGAAGAACGGGATCGCACTTACTCTTTCGCGGGAACAAT 1884  
 447 .....LeuValGlyAlaSerAlaasn 453  
 1885 TTAACGGCAATATACAG.....CAAC 1907  
 454 IleasnAsnAsnLeuSerValLysSerGlyAlaLysPheLysAlaGluThr 470  
 1908 AAACGGCAACACTGTTTTCACGCGCAGCACCGCACCGCCTCAAT 1956  
 470 rAsnspAsnLeuasnIleThrGlyThrPheThrAsnAsnGlyThrSeri 487  
 1957 .....CAATTAGGAAGCGGTGTCAAAANTGAAGATATCCACAGA 2001  
 487 leIleaspValLysLysGlyAlaAlaLysLeu.....Gly 498  
 2002 GAATTCGTGTGGACACAGATTGATCGACCGCACATTAAAGCGGAAA 2051  
 499 AsnIleThrAsnAspGlyAsn...LeuasnIleThrThrAsnAlaLysAs 514  
 2052 CTTCCATATTCAGCGCGACACAGCGGTTCCTCCGCAATGTGCCAAAG 2101  
 514 n.....GlyGlnLysSerValIleasnGlyasnIleThrAsnA 527  
 2102 TGGAAGCGCATTTGCAATTAAGCATTCAGGCCAA.....GCA 2139  
 527 snLysGlyAlaLeuasnIleThrAsnAsnGlyAsnAspThrGluIleGln 543  
 2140 GTTTTCGGTGTGCAACCGCATCAAGCCACACAACTCTACACGTTCGA 2189  
 544 IleGlyGlyAsnIleSerGlnLysGluGlyasnLeuThrIleSerSera 560  
 2190 CTGAGAGGGTCTGCAAGTTTACGGA...AAAACATTAACGAC... 2232  
 560 pLysIleasnIleThrLysArgIleGluIleLysAlaGlyThrAspGln 577  
 2233 .....GATAAGTGAATGCTTCATTCATGACAAACCGACATCGAGAGC 2274  
 577 LysAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThrIleLys... 592  
 2275 AATGTACGCTTCGCGATCAGCTCATTTAATCTACAGCA..... 2316  
 593 ThrLysGlnLeuLysLeuThrGluAsnLeuasnIleSerGlyPheAspLys 609  
 2317 .....CTTGCACACTCAACGGCAATCTTAAGTCAGAGC..... 2349  
 609 sAlaGluIleValAlaLysGluAsnAsnLeuLeuIleGlyAsnAsnA 626



```

2350 ..GAGAC.....ACGCACTATACGCTTACGCGACGACCCAGCAAC 2391
      |||||      ::      |||||      |||      :::::
626  snGlyAspSnaIaAsnAlaValThrValThrPheAsnAsnValLysasp 642
      :::::
2392  GGCACACCTCAGCCCTGCGGCAATGCCAAGCAACATTATCAAGCCAC 2441
      :::::
643  SerLysIleSerAlaasnGly.....HisAsnValThr 653
      |||
2442  ATTAACGGCAACACATCGGCTTCGAC.....AATGCTCATTTA 2482
      |||||      :::::
653  rIeuAsnSerLysValGluThrSerAspGlyAsnSerAsnThrGluGly 670
      :::::
2483  ATCTAAGCAACACGCC.....GTA 2502
      |||
670  snSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVal 686
      :::::
2503  CAAACGGC.....AGTCGACGCTTCCGAC.. 2529
      :::::
687  AsnAsnAspIleThrSerHisLysThrValAsnIleThrAlaSerGluAr 703
      :::::
2530  ....AACGCTAAGCAACGTAAGCAATTCGCACTCAAGCGCAATGCT 2575
      :::::
703  gIleAspThrLysAlaAspThrThrIleAsnAlaThrThrGlyAsnValL 720
      :::::
2576  CCTTAGCCGATTAAGGCACTATTCATTGTAACAGCCGCTTACCGGA 2625
      :::::
720  ySleu.....ThrAlaValThrSer 726
      :::::
2626  AAAATCAGCGCGCGCAAGATACGCAATTAACCTTACAGCAAGCA 2725
      |||||
727  AspIleGlnGlyIleLysSer.....AsnSerGlyAspValAsnI 741
      :::::
2676  GACGCTGCGCTCGGCGACGCAATTAACCTTACAGCAAGCA 2725
      |||
741  eThrThrSerThrGlySerIleAsnGlyLysIleGlnSerLys 758
      :::::
2726  CCATTACACTCAATTCGCGCTTACGACAGTGCGGCGCGCAAC 2775
      :::::
758  eValThrLeuThrAlaThrGluLysThrLeuThrValGlyAsnValSer 774
      :::::
2776  GGCAGTGGCGAGATGGCGCGCGCGCTTCGCGCTTCCATTAATC 2825
      |||||
775  GlyAsnThr.....ValThr 779
      :::::
2826  CGTTACGCGCGCAACTTCGCGAGATCCGCTTCAACACGCGTACGCTAA 2875
      ::|||
779  rValThrAlaAsnArgIleAlaLeuThrThrLeuAlaGlySerThrIle 796
      :::::
2876  ACGGCAATTGAACGCTCAGGAAACATTCGCTTATGTCGAACCTTC 2925
      |||||
796  snGly...ThrAsnGlyValThrThrSerSerGlnSerGlyIleGly 811
      :::::
2926  GGCACCGCGCGCAATTCGCGAGATTCGCGAAAGTCCGCAAGGCACTTA 2975
      |||
812  GlnValValThrGlyLysThrValSerValThrAlaThrLaglySerLe 828
      :::::
2976  CACCTTGGCTCAACAATACCGCGCAAGAACCGTAAGTCTGAGCAAT 3025
      |||||
828  uThrValLysGlyLysIleAsnAlaThrGlnGlyThrAlaThrL 845
      :::::
3026  TGACGAGTAGTGAAGAAAGACACACACCGCTGTCGAAATCTTAAT 3075
      |||||
845  eutThrAlaSerSerGlyLysLeuThrThrGluAlaSerSerAsnIleThr 861
      :::::
3076  TTCACCTTGCAAAACGACAGTCGATCGCGCATCGGCTTATGACGT 3125
      |||
862  Ser.....AlaLysGlyGlnValAspL 869
      :::::
3126  TATCCGAAAGACGCGAGTTCGCGCTGCATTAATCCGGTCAAGAACAG 3175
      |||||
869  eutSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAsnVal 885
      :::::
3176  AGCTTCCGACAAACTCGGC.....AA 3197

```

```

      :::::
886  ThrIeuAsnThrThrGlyThrLeuThrThrValGluGlySerSerIleAs 902
      :::::
3198  GCGGGAAGAAACAGAGCCGCTTGACGCAAAACAGGCAACTTGCCG 3247
      |||||
902  nAlaAsnGluGlyThrLeuValIleAsnAlaAsnAspAlaLysLeuAsp 919
      :::::
3248  CCAACACACAGCGCGGAAAAAGACACACCGCAAGCCCTGACGCGCTGAT 3297
      :::::
919  LysAlaSerGlyAsnArgThrGluValAsnAlaThrAsnAla..... 933
      :::::
3298  GCGCGCGCGCATGCCACCAAGCAAGT..... 3333
      :::::
934  SerGlySerGlySerValThrAlaLysThrSerSerValAsnIleThr 950
      :::::
3334  .....GTTGCCGAACCGCGCC 3349
      :::::
950  rGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluAsnGly 967
      :::::
3350  GGCAGGCA.....GCGGGGA.....AAT 3369
      |||||
967  rGAsnThrValArgLeuArgGlyLysGluIleGluValLysTyrIleGln 983
      :::::
3370  GCGGCACTTATCGACGCGAGGAA.....GACAAAAACGGGTG..CA 3410
      |||||
984  ProGlyValAlaSerValGluGluValIleGluAlaLysArgValLeuG 1000
      :::::
3411  GCGCGATTAAGACACCGCC.....TTGGCGAA..... 3438
      |||||
1000  uLysValLysAspLeuSerAspGluGluArgGluThrLeuAlaLysLeu 1017
      :::::
3438  ..... 3438
      :::::
1017  LyValSerAlaValArgPheIleGluProAsnAsnThrIleThrValAsn 1033
      :::::
3439  ....CAGCGCAAGCGGAACCGCGCGCTTAC 3468
      |||||
1034  ThrGluAsnGlnLupherThrThrArgProSerSer 1044
      :::::
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA01836
seq_documentation_block:
ID   AAB01836 standard; Protein, 1079 AA.
XX
AC   AAB01836;
XX
DT   11-SEP-2000 (first entry)
XX
DE   Haemophilus influenzae strain CDC2 HMW2A protein, SEQ ID NO:47.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-lysable Haemophilus influenzae; NTHI; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain CDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PE   07-OCT-1999; 99WO-CAN00938.
XX
PR   07-OCT-1998; 98US-0167568.
XX
PR   08-DEC-1998; 98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM, Yang Y, Klein MH;
XX
WP   2000-303789/26.
XX
DR   N-PSDB: AAA52185.

```

XX Nucleic acid molecule for producing recombinant high molecular weight  
 PT proteins of Haemophilus which are used as a vaccine to provide  
 PT protection against Haemophilus induced diseases in humans -

PS Claim 12: Fig 23A-P; 307P; English.

XX The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.  
 CC influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene  
 CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,  
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins  
 CC and the hmwB and hmwC genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMWA  
 CC proteins. The modified hmwABC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwA genes (AA52175-A52198)  
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae  
 CC strains Joyce, K1, K21, LDCD2, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or,  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMWA protein from a non-typeable strain of  
 CC H. influenzae.

XX Sequence 1079 AA:

alignment\_scores: Quality: 269.50 Length: 1095.  
 Ratio: 0.496 Gaps: 54  
 Percent Similarity: 49.589 Percent Identity: 20.639

alignment\_block: US-09-303-518D-653 x AAB01836 ..

Align seg 1/1 to: AAB01836 from: 1 to: 1079

```

787 TCATTGGCGACAGTGGCTACACATGTTTATCATGATGCCCAAGCA 836
      ::::::::::::::::::::
78  AsnIleGlyAspSerGlyHis...LeuThrLeuTyrLysArgLysAs 93
      ::::::::::::::::::::
837 AAGGTGTAATTAATGGGTATTGCAAGCAAGCAACCCATATAGGAA 886
      ::::::::::::::::::::
93  nArg.....SerAspGlyIleGlnIleAsnLysAspIleThrSert 107
      ::::::::::::::::::::
887 AAGCAATGCGTCCAGCTAGTGGTAAAGATTGG.....TTCATGAT 930
      ::::::::::::::::::::
107 hngIyGlySerLeuThrIleAsnSerAspTyrValAspIleHis 123
      ::::::::::::::::::::
931 GAATCTTGGCTGAGATACCATTCAGTATTCAGACCAACATCAAAA 980
      ::::::::::::::::::::
124 AsnIleThrLeuGlyGly.....PheLeuAsnIleThrSerSe 137
      ::::::::::::::::::::
981 TGGCAAACTATTTTTAAGCAATATATAGCGCAGAAATTCATG 1030
      ::::::::::::::::::::
137 rAspSerValAlaPheGlyGlyAsnGlyAsnLysArgSerSera 154
      ::::::::::::::::::::
1031 CCAAAACATAAA.....CACTATCTCTACTTATAGATTAATAACA 1071
      ::::::::::::::::::::
154 lAserAlaGlnIleIleAlaGlnGlyThrIleThrGlyGluAsn 170
      ::::::::::::::::::::
1072 CGAACCGTTCATTTTATGTTTCTTTATCCGAGACAGCAAGAAC 1121
      ::::::::::::::::::::

```

```

171  LysThrPheArgLeuAsnAsnValSerLeuAsnGlyThr..... 183
1122 TGTATTATCATGCTGCAGGTGGGGTTCACAGTATTCAGCCAGCTGATA 1171
183 ..... 183
1172 AVGAGAAATATTTCTTTATTGACAAAGAAAGTGAATGATACTT 1221
      ::::::::::::::::::::
184 ..GlyAsnGlyLeuSerIleIleSer.....Thr 192
1222 ACCAGCAATCATACCAAGCGCGCGCT...TTCATTTTGAAGCTTA 1268
      ::::::::::::::::::::
193 AlaSerAsnLeuSerHisArgLeuAspGlyIleIleAsnValSerGlyAs 209
1269 TTTTACGTC.....TCGCTTAAACAAACGAAACGTGGCAAGGC. 1308
      ::::::::::::::::::::
209 nValThrIleAsnGlnThrThrGlnIleAsnIleGlyThrTrpLysAlas 226
1308 ..... 1308
226 eSerAspSerTyrTrpAsnValThrSerPheAsnLeuArgGluAspSer 242
1308 ..... 1308
243 LysPheThrPheIleLysTyrValAsnSerAlaArgAsnGlyAspValAr 259
1309 .....GCGGGCGTT..... 1317
259 gGlyArgSerPheAlaGlyValIlePheAsnAlaLysGlyLeuThrTrs 276
1318 .....CATTCAGTGTGTCAGTACCGTTACTTGGAAAGTA.....AAC 1356
      ::::::::::::::::::::
276 ePheAsnValLysLysGlySerThrValAspPheLysLeuLysProAsn 292
1357 GCGGTGGCAACGACCGC.....CT 1376
      ::::::::::::::::::::
293 SerGlyTyrAsnSerGlnIleArgSerSerIleAsnValSerAspGly 309
1377 GTCCAAATATCGGCMAAGCAGCGCTGCTGTTCAAGCCAAAGGAAACC 1426
      ::::::::::::::::::::
309 eSerValSerGlyGlyArgValAsnIleAsnThrLeuAlaIleAsnLeu 326
1427 AAGGC.....TCGTCAGCGTGGCGACGGT 1452
      ::::::::::::::::::::
326 hngIyGlyGlyValGluIleArgSerSerIleAsnValSerAspGly 342
1453 AAGTCATC...TTAGATCAGAGCGGACGATCAAGCAAAACAAAC 1499
      ::::::::::::::::::::
343 SerThrLeuSerMetThrAlaGlnAlaArgAspArgAsnAlaPheGlu 359
1500 CTTTAGTGAATCGGCTTGGTCAAGCGGACGGGACGGTCAACTGAATG 1549
      ::::::::::::::::::::
359 eThrLysAspLeuValIleAsnAlaSerAsnSerAsnLeuSerIleG 376
1550 CGGATATACGTTCAACCCCGCAACACTCTATTTCGCTTTCGGCGGCA 1599
      ::::::::::::::::::::
376 hngIleAsnAsp.....GlyPhe..... 381
1600 CGTTGGATTGGAACGGCATTCGTTTGGTTCACCGCATTCAAAT.. 1647
      ::::::::::::::::::::
382 .....AspAsnAsnGlnLysAlaAsnAlaIleAsnSerLysTyrAsnVa 396
1648 ACCGATGAAGGGCGATGATGTCACACACATCAAGCAAAAGATCCA 1696
      ::::::::::::::::::::
396 lThrIleGlnGlyGlyAsnValThrLeuGlyGlnAsnSerSerSert 413
1697 CGGTTACG.....ATTACGCAATTAACATATTTACTACAAAC 1734
      ::::::::::::::::::::
413 hrlleThrGlySerValAsnIleGlyAlaAsnAlaAsnValThrLeuGln 429
1735 GCGAATTAACAACTTGATACCAAAAAAGAAATTCCTACAAACGGTTG 1784
      ::::::::::::::::::::
430 AlaHisAsnGlyAsn...AspArgAsnLysLysLeuThrPheGlyAsn.. 444

```

1785 GTTGGCGAGAAAGATGCACCAAAAGACGGGGCTCATCTGATTT 1834  
445 .....ValSerValGluGlyGluLeuArg..... 452  
1835 ACCAACCAGAAAGACGGATGCATTTACTGCTTTCCGGGCAACAAT 1884  
453 .....LeuValGlyAlaSerAlaAsn 459  
1885 TTAAACGGCAATATCACG.....CAAAC 1907  
460 ILeuAsnAsnLeuSerValLysSerGlyAlaLysPheLysAlaGluTh 476  
1908 AAACGGCAAACTGTTTTCAGCGCAGACCGACCGACCGCACAT. 1956  
476 rAsnAspAsnLeuAsnIleThrGlyThrPheThrAsnAsnIleThrSeri 493  
1957 .....CATTTAGAAAGCGGGTGTCAAAATGAAAGGTATCCCAAGCA 2001  
493 leIleAspValLysGlyAlaAlaLysLeu.....Gly 504  
2002 GAAATCGTGTGGACAAAGATGTGATGACCGCACATTTAAAGCGGAAA 2051  
505 AsnIleThrAsnAspGlyAsn..LeuAsnIleThrThrAsnAlaLysAs 520  
2052 CTTCATATTCAGGCGGACAGCGGTGTTCGCCGCAATGTTGCCAAG 2101  
520 n.....GlyGlnLysSerValIleAsnGlyAsnIleThrAsnA 533  
2102 TGGAAGCGCATGTGCATTTAAGCATACGCCCAA.....GCA 2139  
533 snLysGlyAlaLeuAsnIleThrAsnAsnGlyAsnAspThrGluIleGln 549  
2140 GTTTTCGGTGTGCACCGCATCAAGCCACACATCTGTACAGTTGGA 2189  
550 IleGlyLysAsnIleSerGlnLysGlnLysAsnLeuThrIleSerSerAs 566  
2190 CTGAGCGGGTCTGACAAGTTGTACGAA...AAACCATTTACGAC... 2232  
566 pLysIleAsnIleThrLysArgIleGluIleLysAlaGlyThrAspGln 583  
2233 .....GATTAAGTATGCTTCATTTGAGCAAGACCGCATTCAGGCG 2274  
583 LysAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThrIleLys... 598  
2275 AATGTACGCTTGGCGATCACGCTCATTTAATCTCACAGA..... 2316  
599 ThrLysGluLeuLysLeuThrGluAsnLeuAsnIleSerGlyPheAspLys 615  
2317 .....CTTGCCACACTCAACGGCATCTTAAGTCAGGC..... 2349  
615 sAlaGluIleValAlaLysGluAsnAsnLeuIleIleGlyAsnAsnA 632  
2350 ..GGAGAC.....ACGCACTATACGGTTACGGCAAGCCACCGCAAAAC 2391  
632 snGlyAspAsnAlaAsnAlaLysThrValThrPheAsnAsnValLysAsp 648  
2392 GCGAACCCTGAGCTGTGGCAATGCCCAAGCAACATTTATCAACGCC 2441  
649 SerLysIleSerAlaAsnGly.....HisAsnValTh 659  
2442 ATTAACGGCAACATCGGCTTCGAC.....AATGCTTCATTTA 2482  
659 rLeuAsnSerLysValGluThrSerAspGlyAsnSerAsnThrGluGlyA 676  
2483 ATCTAAGCACAAAGCC.....GTA 2502  
676 snSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVal 692  
2503 CAAAGCGC.....AGTTCAGCGCTTTCGAC... 2529  
693 AsnAsnAspIleThrSerHisLysThrValAsnIleThrAlaSerGluThr 709  
2530 ....AAGCTTAAGCAAACTAAGCCATTCCGCACTCAACGGCAATGCT 2575  
709 gIleAspThrLysAlaAspThrThrIleAsnAlaThrThrGlyAsnValTh 726  
2576 CCGTAAGCCGATTAAGCAGTATTCATTTGAAACAGCGCTTACCGGA 2625  
726 LysLeu.....ThrAlaValThrSer 732  
2626 AAAATCAGCGCGGCAAGGATACGCGATTAACCTTAAAGACAGCGAATG 2675  
733 AspIleGlnGlyLysIleLysSer.....AsnSerGlyAspValAsnIle 747  
2676 GAGCTGCGCGTGGCGACGCAATTAAGCAATTTAACTTGACACAGCCA 2725  
747 eThrThrSerThrGlySerIleAsnGlyLysIleGluSerLysSerGlyS 764  
2726 CCATTAACACTCAATTCGCGTATCGACAGATGCGGCGCGGCAAC 2775  
764 eValThrIleuThrAlaThrGluLysThrLeuThrValGlyLysValSer 780  
2776 GGCATGCGCGCAGATGCGCGCGCGCGCGCGCGCTTCCTATATC 2825  
781 GlyAsnThr.....ValTh 785  
2826 CGTTACGCGCGCACTTCGCGAGAAATCCCGTTTCAACAGCGTACGAT 2875  
785 rValThrAlaAsnArgIleAlaLeuThrThrIleuAlaGlySerThrIleA 802  
2876 ACGGCAAAATGAAAGCGGTGAGGACATTCGCGTTATGCGCAACTCTTC 2925  
802 snGly...ThrAsnGlyValThrThrSerSerGlnSerGlyIleGly 817  
2926 GCGTACCGCAGCGGCAATTTGAAGCTGCGGCAAGTCCGAGACACTTA 2975  
818 GlyGluValThrGlyLysThrValSerValThrAlaThrAlaGlySerLe 834  
2976 CACCTGGCTGTCAACCAATACCGGACAGAACCCGATGTCGAGCAT 3025  
834 uThrValLysGlyGlyAlaLysIleAsnAlaThrGluGlyThrAlaThrL 851  
3026 TGACGTAAGTGAAGAAAGACACACACCGCTTCGGAATCTTAAT 3075  
851 eThrAlaSerSerGlyLysLeuThrThrGluAlaSerSerAsnIleThr 867  
3076 TTCACCGCTCAAAAGACACGCTGCATGCGCGCATGCGCTTATCACT 3125  
868 .Ser.....AlaLysGlyGlnValAspL 875  
3126 TATCGCAAGACGCGAGGATTCGCGCTGATTAATCGGTCAAAGCAAG 3175  
875 eSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAsnVal 891  
3176 ACCTTTCGACAACCTGCG.....AA 3197  
892 ThrLeuAsnThrThrGlyThrLeuThrThrValGluGlySerSerIleAs 908  
3198 GCGGGGAGAAACAGAGCGCGCTGACGGCAAAACAGCACAACTTGCGG 3247  
908 nAlaAsnGlnGlyThrLeuValIleAsnAlaAsnAlaLysLeuAspG 925  
3248 CCAAAACAGCGCGGAAAGACAGCGCAAGCGCTTACGCGCTGATTT 3297  
925 LysAlaSerGlyAsnArgThrGluValAsnAlaThrAsnAla..... 939  
3298 GCGCGCGGCGCAATGCAACCGAAAGCAGCAAGT..... 3333  
940 SerGlySerGlySerValThrAlaLysThrSerSerSerValAsnIleTh 956  
3334 .....GTTGCGAAGCGCGCC 3349  
956 rGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluAsnGlyA 973  
3350 GCGCAGCA.....GCGCGGCA...AAT 3369



563 euglAlaInglyAsnIleAsnIleThralAlaIySlnsPlleaIaphe 579  
637 AACCGCAAAAGTTCATATCATATATTCGAAGCGCATATCTTGCGCTCGCG 686  
580 GluIySglSerAsnGlnValIle..... 587  
687 TGGCAATACCTTGGACAAAAATGGATCAGAGTGTGGCACAGTCAACTTG 736  
588 .....ThrgIyGlnGlyThrIleThrsGrg 596  
737 GTAGCAAAAAATTAAACATAGCCCATATGTTTATTACCAAGAGGAGCG 786  
596 LysAsnGlnIys.....GlyPheargPheAsnSsnVal 606  
787 TCATTTGGGCGACAGTGGGTCACCAATGTTTATCTATGATGCCCAAAAGCA 836  
607 SerLeuAsnSlyThrIlySerGlyLeuGlnPheThrIlyArgThrAs 623  
837 AAGGTGTTAATTAAATGAGGTATTCGAACACAGCAACCCCTATATAGGAA 886  
623 nIySTyAlaIleThrsAsnIySphleuIyThrIleAsnIleSergIyL 640  
887 AAGCAAT..GGCTTCAGAGCTAGTTCGTAAGAATGGTTCTATGATGAA 933  
640 ySAlAsnIleSerMetValIleuProIyAsnSlnSergIyTyAspIys 656  
934 ATCTTGTGAGATATCCCAT.....TCAGATCTTACAGACCAACACa 977  
657 ..PheIySglYArgThrIyThrIyPheAsnIleuThrsIleAsnValSergI 672  
978 AATGTGGAATACTTTTAAACGACATATATATGCG.....CGAG 1018  
672 uSerGlyGlnPheAsnIleuThrIleAspSerArgIySerAspSeraIag 689  
1019 GAAAAATCGATGCGCAACATTAACATATATCTCTACCTATATGATTAA 1068  
689 LyrhIreuthrGlnProIyAsnIleAsnGlyIleSerPheAsnIySasp 705  
1069 ACAGCAACCGTTCATATGTTATATGTTCTTATTCGAGACAGCAAGAA 1118  
706 Thr.....ThrsPheAsnValGlnuAsnAlaArgValAsnph 718  
1119 ACCTGTTTACAGTCGCGAGGGGGTCAACAGTATAGACCCAGACTGA 1166  
718 eAspIle...LysAlaProIleGlyIleAsnIySthr...SerSerIeua 733  
1169 AT.....ATAGGAGAAATATTTCTCTTATTCAGCAAAAGAAA 1206  
733 snIyTrAlaSerPheAsnGly...AsnIleSerValSergIyGlyIySerg 748  
1207 GGTGATGTG..ATCTTACCGCAACATCAACCAAGCGCGGGGGGTTT 1255  
749 ValAspPheThrIleuAsnIleAsnIleSerSerAsnValGlnThrProGlyIy 765  
1254 GTATTTTGAAGGTAATTTACGCTCGCGCTAAAACAAGCAAAAGTGGC 1303  
765 ValIleAsnSerIySthr..... 771  
1304 AAGCGCGGGCGCTCATATCACTGATGCAGTACCGCTTACTTATTCGAAGTA 1353  
772 .....PheAsnValSerThrGlySerSerLeuArgPheIySthr 784  
1354 AAGCGC.....GGGCAAAAGCGC.....CGCGTCTC 1379  
785 SerGlySerThrIlySphleuIySergIleGlyIySAspPheThrIleAs 801  
1380 CAATAATGCCAAAGCAGCGCTGCTGTTCAACCAAGAGGAGAAACAA 1428  
801 nAlaThrGlyGlyAsnIleThrsIleuIleuGlnValGlnIyThrAspGly 818  
1429 .....GGCTCGGCAAGCGCGCGCGCAAGCAAGTATCTTACATCAGAG 1473  
818 etIleGlySglIyValAlaIySAsnIleThrsPheGlnIyGly 834

[illegible]

```

2323 ACACCTAAGCGCAATCTTAGTGCAGGCGGAGACGACGACATATACGGTTAC 2372
1062 ThrIleValThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078
2373 GCGCAAGCGCCCAAAACGCAACCTCAGCCTCGTGCGCAATGCCAAG 2422
1078 rPhenGlnValLysAspSerLysIleSerAlaAspGlyHisLysValT 1095
2423 CAACTATTATCAAGCCACATTAAAGCGCAACATCGGCTTGGAACAT 2472
1095 hrLeuHisSerLysValGluThrSerLysAsnAsnValThrGluAsp 1111
2473 GCTCATTTATCTAAAGCAACAGCGGTACAAACGCGCTGTCGCGCT 2522
1112 SerSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVa 1128
2523 TTCGACACAGCGTAAGGCAAAAC.....GTACCCATTTCCGACCTAACG 2566
1128 LAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSerG 1145
2567 GCATATGTC.....TTCCTAGCCGATAGGCAATATTCAT 2601
1145 LysIleIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAsn 1161
2602 TTTCAAACAGCGCGCTTACCGGAAATACGCGGCGCAAGGATACG.. 2649
1162 ValGluIleThrAlaGlnThrGlySerIleLeuGlyLysIleGlnSer 1178
2650 .....GCATTACCTTAAAGACAGCAAGATGAGCTCGCTCGGGCA 2692
1178 rSerGlySerValThrLeuThrAlaThrGluGlyAlaLeu..... 1191
2693 CGGAATTAGCAATTAAACCTTGACAAACGCCACATTAACATCAATTC 2742
1192 ..AlaValSerAsnIleSerGlyAsnThrValThrValThrAlaAsnSer 1207
2743 GCCTATGACACAGATGCGGAGCGCGCAACCGCGACGCGAGATGC 2792
1208 .....GlyAlaLeuThrThrLeuAlaGlySer.. 1216
2793 GCGCGCGCGCGCTCGCGCGCTCCATTATCCGTTACCGCGCACTT 2842
1217 .....ThrIleLysGlyThrGluSerValThr.....Thrs 1227
2843 CGGCAAGATCCCGTTTCAACACGCTGACGAAACGCGCAATTGACGT 2892
1227 erSerGlnSerGly.....AspIleGlyGlyThrIleSerGly 1239
2893 CAGGGAACA.....TTCGCTTATGTCGGAACCTTCGCGTACGCGCAG 2936
1240 ..GlyThrValGluValLysAlaThrGluSerLeuThrThrGlnSerAs 1255
2937 CGGCAATTTAGCTGCGGGAAGTTCGGAAGGACTTACACCTTGCGCTG 2986
1255 nSerLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAlaT 1271
2987 TC.....AACATACCGCGCAACGCAACGCGTA 3012
1271 hrGlyThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAla 1287
3013 AGTCTCGACAAATTGACGGTA.....GTGGA 3038
1288 AsnAlaGlyAspLeuThrValGlyAsnGlyValaGluIleAsnAlaThrG 1304
3039 AGGAAGAAGACACACACCGCTGTCGGAATCTTAATTTCACCTCGCAAA 3088
1304 uGlyValAlaIleThrLeuThrThrSerSerGlyLysLeuThrThrGluAla 1321
3089 ACGAAGATCGATGCGCGCATGCGCTATTCAGTTATCCGCAAGAC 3138
1321 erSerHisIleThrSer..AlaLysGlyGlnValAsnLeuSerAlaGlnAs 1337
3139 GCGCAGTTCCGCTGCATATATCGGTCAAGACAAAGAGCTTTCGACA 3187

```

```

1337 polYserValAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353
seq_name: /SUS1/9cdata/geneseq/geneseqp-emb1/AA2000.DMT:AA201846
seq_documentation_block:
ID   AAB01846 standard; Protein; 1536 AA.
XX
AC   AAB01846;
XX
DE   11-SEP-2000 (first entry)
XX
DE   Haemophilus influenzae strain 12 HMW1A protein, SEQ ID NO:67.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   W0200020609-A2.
XX
PD   13-APR-2000.
XX
PE   07-OCT-1999; 99WC-CA00938.
XX
PR   07-OCT-1998; 98US-0167568.
PR   08-DEC-1998; 98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM, Yang Y, Klein MH;
XX   WPI: 2000-303789/26.
XX   N-PSDB: AAA52195.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Example 16; Fig 28A-Q; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae
CC   strains Joyce, KI, K21, ICC02, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typable strain of
CC   H. influenzae.
XX
SQ   Sequence 1536 AA;

```

alignment\_scores:

Quality:	268.50	Length:	1217
Ratio:	0.465	Gaps:	60
Percent Similarity:	47.494	Percent Identity:	20.624

```
alignment_block:
US-09-303-518D-653 x AAB01846 .
```

Align seg 1/1 to: AAB01846 from: 1 to: 1536

```

1 ATGAAACAAACCCGACAAACGAGAACCGACAAACCGAACCGCAAGCCCTTA 50
327 MetIethrNglyAspIysValThrLeuLysThrGlyAlaValIleaspLe 34.34
51 AACCGGCCGCAATCCGCGCTTCTGCCCCGCTTACTTAGCCATAGCCTGCTG 100
343 uSerLysLys..... 34.68
101 TCGCAATTCGCCCCCAAGCCCGGGGAGACACTTATTTCCGGCATCAAC 150
347 ..... GluIleGlyGluThrTyrLeuGly..... 35.4
151 TACCAATACTATCGCACTTCTGCCGAATAAAGCAAGTTTGCACTGCG 200
355 ..... GlyAspGluArgIleGlu..... Gl 36.6
201 GGCCAAGATATTAGCTTTATCAACAACAAAAAGCGAGTGTGCGCAAT 250
361 LysLysGlyLysIleGluLeuAlaLys..... ThrS 37.2
251 CGATGACGACAAAGCCCCGATGATATTCTTCTGTGATCGCGTAACGCG 300
372 erLeuGluIysGlySerThrIleAsnValSerGlyLysGluLysGlyGly 38.8
301 GTGGCGCATTTGCGGGCGCATCAATATATTGTG..... 33.3
389 ArgAlaIleValTrpGlyAspIleAlaLeuIleaspLysAsnIleAsnAl 40.5
334 ..... AGCGTGGCACAATAACGGCGCGTAT..... 35.7
405 agInGlySerGlyAspIleAlaLysThrGlyGlyPheValGluThrSerg 42.2
357 ..... 35.7
422 LysHisAspLeuPheIleLysAspAsnAlaIleValAspAlaLysGluTrp 43.8
358 ..... AACATGTGTTGATTTTGGTGGCGAG..... 36.1
439 LeuLeuAspProAspAsnValSerIleAsnAlaGluThrAlaGlyArgSe 45.5
382 ..... GGAAGCAATGCCCATCAGACGCG..... TTTCTTACCAAA 41.8
455 rAsnThrSergGluAspSerLysThrGlySergLysAsnSerAlaSerT 47.2
419 TTGTGAAAAGAAATTAAT..... 43.5
472 hrProLysArgAsnLysGluLysThrThrLeuThrAsnThrThrLeuGlu 48.8
436 ..... TATAAGCAGGAGCT..... 45.0
489 SerIleLeuLysLysGlyThrPheValAsnIleThrAlaAsnGlnArgIl 50.5
450 ..... 45.0
505 eTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpS 52.2
451 ..AACGGCATCTCTTATGGCGGCGATATCATATGCGCGTTTGCAAAA 49.8
522 erGluIysArgSergGlyGlyGly..... ValGluIleAsnAsn 53.4
499 TTGTGTACA..... GATGAGAAACCTGTTGAATGACCACTATAT 53.9
:::||||| ..... 33.3

```

535 AsplIethrThrcIylaspbsprhrArgVAlaAsnLeuThrIleTyrSe 551  
540 GGATGGGTGGAATACCGCGATTTAAATATAACCTGATCGTGTGGA 589  
551 rgllylTTP....ValAspValHisLys.....AsnIleSerL 563  
590 TCGGAGACA...GGCAGACATATATGGCGGTGATGAGACGACACCAAT 636  
563 euclYAlaGlnGlyAsnIleAsnIleThrAlaLysGlnAspIleAlapHe 579  
637 AACCCGGAAAGTTCAATCATATATTCGACAGCATATTCCTGGCTGCG 686  
580 GlulYsgIySerAsnGlnValIle..... 587  
687 TGGCAATACCTTGCACAAATAGATCAGGTGGTGGCAAGTCAACTTGA 736  
588 .....ThrlYglnGlnThrlIethrSerG 596  
737 GTAGCAAAAAATTAAACATAGCCATATAGTTTTTTACCAACAGAGAGC 786  
596 LysAsnGlnLys.....GlyPheArgPheAsnAsnVal 606  
787 TCATTTGGGAGACAGTGGCTACCAACATGTTATCTATGATGCCAAACCA 836  
607 SerLeuAsnGlnIethrLysSerLylEuGlnPheThrLysArgThAs 623  
837 AAGATGGTAAATTAATGAGGATTTGCAACACAGCAACCCCTATATAGAA 886  
623 nlYsYrAlaIleThrAsnLysPheGlnGlyThrlLeuAsnIleSerGlyL 640  
887 AAGCAAT...GGCTTCACGCTAGTTCGTAAGAAGTGTCTATGATGAA 933  
640 YsAlaAsnIleSerMetValIleProLysAsnGlnSerGlyTyrLysP 656  
934 ATCTTGTGAGAGATACCAT....TCAGATTTCTAGCAACCATCA 977  
657 ...PheLysGlyArgThrLylTrpAsnLeuThrSerLeuAsnValSerG 672  
978 AATGGGAATACCTTTTAAACGACATATATATGCG.....CGAG 1018  
672 uSerGlyGlnPheAsnLeuThrIleAspSerArgLysSerAspSerAlaG 689  
1019 GAAAAATCATGCCAAACCTAATACCTATTCCTACCTTATAGATTAA 1068  
689 LylThrLeuThrGlnProTyrAsnLeuAsnGlyLleSerPheAsnLysAsp 705  
1069 ACACGACACCGTTCAATGTTTATATGTTTCTTATTCGAGACAGCAAGA 1118  
706 Thr.....ThrPheAsnValGlnArgAsnAlaArgValAsnPh 718  
1119 ACCTGTTTTCATGCTGCAGCGTGGGGTCAACAGTATGACCCAGACTGA 1168  
718 eaSpIle...LysAlaProIleGlyIleAsnLysTyr...SerSerLeuA 733  
1169 AT.....AATGGAAAAATATTTCTTATATGACAAAGAAAA 1206  
733 snlYAlaSerPheAsnGly...AsnIleSerValSerGlyGlyGlySer 748  
1207 GGTGAATTC...ATATCTACACGACACATCAACCAAGCGCGCGGT 1253  
749 ValAspPheThrLeuLeuIleAsnIleSerSerAsnValGlnThrProGlyVa 765  
1254 GTATTTTGGGTAAATTTTACAGTCTCGCTCAAAAAACAGCAACGAGGC 1303  
765 LylAlleAsnSerLysTyr..... 771  
1304 AAGCGCGGCGCGTTCATATCATGATGCGCACTACCGTTACTTGAAAGTA 1353  
772 .....PheAsnValSerThrGlySerSerLeuArgPheLysThr 784  
1354 AAGCGC.....GTGGCAAAACGAC...CGCTGTC 1379  
785 SerGlySerThrLysThrGlyPheSerIleGlnLysAspLeuThrLeuAs 801

1380 CAAATCGCAAGGACGCTGCTTCAAGCCAAAGGGAACCAA. 1428  
1429 .....GGCTGGTCAGGCGGAGCGTAAAGTCACTTAGATCAGCAG 1473  
801 nAlaThrGlyGlyAsnIleThrLeuGlnValIleGlyThrAspGlyM 818  
818 etIleGlyGlyGlyIleValAlaLysLysAsnIleThrPheGlyGly 834  
1474 GCGAGCATCAGGCAAAAAAAGCCTTTAGTAATCGCTGGTCAAG 1523  
835 AsnIleThrPheGlySerArgLysAlaValThrGluLeu.....G1 848  
1524 GCGAGGAGGAGCGTGAAGTCAATGCGGATATCACTTCAACCCGACA 1573  
848 uGlyAsnValThrIleAsnAsnAlaAsn..... 858  
1574 AACTCTATTCGGCTTCGGCGGAGCGTGGATTGAGAGGCAATCG 1623  
859 .....ValThrLeuIleGlySerAsp 865  
1624 CTTCGTTCCACCGCATTTAAATACCGATGAAGGGCGATGTGCA 1673  
866 PheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleLeas 882  
1674 CCACATCA.....GACAAAGATCCACCGCTTACCATCAGCAATA 1717  
882 nSerGlyAsnLeuThrAlaGlyLysIleValAsnIleIleGlyAsn. 898  
1718 AAGATATTACTACACCGCATATACAACTGTGATAGCAAAAAAGAA 1767  
899 .....LeuThrValIleGlySerAsnAlaAsnPheLysAlaIleThrAsn 912  
1768 ATTCCTTCAAC.....GGTGGTTGGGAGGAAAGATGCACCAAC 1811  
913 PheThrPheAsnValIleGlyLeuPheAspAsnLys..... 924  
1812 GAACGGCGGCTCACTGAATTCACCAAG.....GAA 1846  
925 .....GlyAsnSerAsnIleSerIleAlaLysGlyAlaArgPheLys 940  
1847 AAGCGATCGCATTTACTGCTTCGCGGAGCAAT..... 1884  
940 sPleAspAsnSerLysAsnLeuSerIleThrThrAsnSerSerThr 956  
1885 .....TTAAAGGCAATATCAGCAAGCAAAAGGCAAACTGT 1922  
957 TyrArgThrIleIleSerGlyAsnIleThrAsnLysAsnGlyAspLeuAs 973  
1923 TTTCAGGCGGACGACCGACCGCATCAATCATTTAGAGCGGAGT 1972  
973 nIleThrAsnGlySer...AspThrGluMetGlnIleGlyLysAsp 989  
1973 GGTCAAAATGGAAGGTRATCCACAGAGCAAAATGTTGGACACGAT 2022  
989 aIleSerGlnLysGly..... 994  
2023 TGGATCGACCGCATTTAAAGCGAAACTTCATATTCAGCGGACA 2072  
995 .....AsnLeuThrIleSerSerAspLysIleAsnIleThr..LysG1 1008  
2073 AGCGGTGTTTCCGCAATGTGGCAAGTGAAGCGATTTGCAATTTA 2122  
1008 nIleThrIleLysAlaGlyValAspGlyLysAsnSerAspSerAspAla 1025  
2123 GCATTCAGCCCAAGCAGTTTCGTGTGCGACCGCATCAAGCGACACA 2172  
1025 hTrAsnAlaAsnLeu.....Thr 1031  
2173 ATCTGTACGTTGCGACTGAGCGGCTGTGACAAGTTGTACGGAATAAC 2222  
1032 IleLysThrLys.....GluLeuLys 1038

2223 CATTACGCGCATTAAGTATTGCTTCATGTGAGCAAGACGACATCAGAG 2272  
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluLeu... 1053  
2273 GCAATGTACGCTTGGCCGATCAGCGCTCAATTAATTCACAGGACTTGC 2322  
1054 .....ThrAlaLysAspLysSerAspLeu..... 1061  
2323 ACATCAACGCGCAATCTTAGTCAGAGCGGACAGACGCTATACGTTAC 2372  
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078  
2373 GCGCAACGCCCAACCAAGCGCAACCTCAGCCTGGGCAATGCCAAG 2422  
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGlyHisLysValT 1095  
2423 CAACATTTAATCAACCCACATTTAAAGCGCAACACATCGCTTGGACAT 2472  
1095 hLeuHisSerLysValIleThrSerGlySerAsnAsnAsnThrGluAsp 1111  
2473 GCTCATTTAATCTAAGCAACAGCGCTACAAACGCGAGCTGACGCT 2522  
1112 SerSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVa 1128  
2523 TTCGCAACAGCTTAAGGCAAC.....GTAGCCATTCCGCACTCAAG 2566  
1128 lAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSerG 1145  
2567 GCATATTC.....TCCCTAAGCCGATTAAGCAAGTATTCAT 2601  
1145 lGlyIleThrThrLysThrGlyThrThrIleAsnAlaThrThrLysn 1161  
2602 TTTGAAACAGCGCTTACCGGAAATTCAGCGCGGCAAGATACG... 2649  
1162 ValGluIleThrAlaGlnThrGlySerIleLeuGlyIleGlyLysSer 1178  
2650 .....GCATTACACTTAAGACAGCAATGACGCTGCCGTCGGGCA 2692  
1178 rSerGlySerValThrLeuThrAlaThrGluIleValLeu..... 1191  
2693 CGGAATTACGCAATTTAACTTGACAAAGCGCACCATCACTCAATTC 2742  
1192 ..AlaValSerAsnIleSerGlyAsnThrValThrValThrAlaAsnSer 1207  
2743 GCTATTCAGACAGATCGGACGCGCAACCGGACGTCGCGAGATGC 2792  
1208 .....GlyAlaLeuThrThrLeuAlaGlySer.. 1216  
2793 GCGCGCGCGCGCTTCGCGGCTTCCCTATATTCGCTTACGCCCACTT 2842  
1217 .....ThrIleLysGlyThrGluSerValThr.....ThrS 1227  
2843 CGGCAATCCGCTTCAACACGCTGACGCTAAGCGCAATTTGAAGCT 2892  
1227 eIleSerIleThr.....AspIleGlyGlyThrIleSerGly 1239  
2893 CAGGACA.....TTCGCTTATATGCGAATCTTCGCTTACCGCAG 2936  
1240 ...GlyThrValGluValLysAlaThrGluSerLeuThrThrGluSerAs 1255  
2937 CGGCAATTTGAAGCTGCGGAAAGTTCGAGAGGCACTTACACTTGGCTG 2986  
1255 nSerLysIleLys...AlaThrThrGlyGluAlaAsnValThrSerAla 1271  
2987 TC.....ACAAATACCGGCAACGACCGCTA 3012  
1271 hGlyThrIleGlyGlyThrIleSerGlyAsnThrValAsnValThrAla 1287  
3013 AGTCTGAGCAATTTGAGGTA.....GTGA 3038  
1288 AsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrG1 1304  
3039 AGCAAAAGACACACACCGCTGTCCGAAATCTTAATTCACCGTCGAAA 3088







```
3048 AsnGlnYssSerAlaIleSerSerGlnIleAspArgAlaGlyHisValSe 3064
|||||
3258 GCACATATACGGTTACGGCAACGCC.....ACCCAAACG 2392
|||||
3064 rGluValThrAlaIleLysAsnAlaIleThrGluLeuAsnThrGlnMetG 3081
|||||
2393 GCACATCAGCGCTGTGGCCAAATGCCAGAACATTATTAATCAAGCCACA 2442
|||||
3081 LysAsnLeuGluGlnAlaIleHisAspGlnAsnThrValLysGlnGlyVal 3097
|||||
2443 TTAAACGGCAACACATCGCGCTTCGGACAAATCTTCAATTAATCAACAA 2492
|||||
3098 .....AsnThrAspAlaAspLysAlaLysArgAspAlaTyrThr 3111
|||||
2493 CAAGCGCGTACAAACGGCACTGACGCTTTCGGACACGCTAAGCA 2542
|||||
3111 rAsnAlaValSerArgAlaGluThrIleLeuAsnLysThrGlnGlyAla 3128
|||||
2543 ACCTAAGC..... 2550
|||||
3128 snThrSerLysGlnAspValGluAlaIleGlnAsnValThrSerAla 3144
|||||
2551 CATTCGCCACTCAACGCC.....AATGTCCTCCAGCGATTAAGCAGT 2594
|||||
3145 LysAsnAlaLeuAsnGlyAspGlnAsnValThrAsnAlaLysAsnThrAl 3161
|||||
2595 ATTCCATTGTTGAAACAGC..... 2614
|||||
3161 aLysHisAlaLeuAsnAsnLeuThrSerIleAsnAsnAlaGlnLysArg 3178
|||||
2615 GCTTTACCGGAAAAAATCAGCGCGGCAAGATACGCGATTACACTTA 2664
|||||
3178 spLeuThrThrLysIle.....AspGlnAlaThrThrValAla 3190
|||||
2665 GACAGCGCAATGAGCGCTGCGCGGCAAGCAATTAAGCAAT.....TT 2708
|||||
3191 GlValGluAlaValSerAsnThrGlyThrGlnLeuAsnThrAlaMetAl 3207
|||||
2709 AAACCTTGACAAAGCGC.....AACATTACACTCAATTCGCGCT 2746
|||||
3207 aAsnLeuGlnAsnGlyIleAsnAspLysAlaAsnThrLeuAlaSerGlu 3224
|||||
2747 ATGCACACGATGCGGCGCAGGC.....CGCAAAACCGGCGAGCG 2784
|||||
3224 snTyrHisAspAlaAspSerAspLysLysThrAlaTyrThrGlnAlaVal 3240
|||||
2785 GCAGATCGCGCGCGCGCGCTGCGCGCGCTCCCTATTATCGCTTACGCC 2834
|||||
3241 ThrAsnAlaGluAsnIleLeuAsnLysAsnSer...GlySerAsnLeuAs 3256
|||||
2835 GCCAACTTCGGCAAGATCCGCTTCAACACGCTGACG..... 2871
|||||
3256 pLysAlaIleValAlaGlnAsnAlaLeuSerGlnValThrAsnAlaLysGly 3273
|||||
2872 ..GTAAACGCC.....AATTTGAACGGTCAAGGA 2898
|||||
3273 LaleuAsnGlnLysHisAsnLeuGlnAlaIleLysSerAsnAlaAsnThr 3289
|||||
2899 ACATTCGCGTTTATGTCGGAACCTTTCGCTACCGCGACGCGCAATTGAA 2948
|||||
3290 ThrIleAsnGlyLeuGlnHisLeuThrThrAlaGlnLysAspLysLeuL 3306
|||||
2949 G.....CTGGCGAAAGTTCCGAAGGCACTTACACCTTGCGCTG 2986
|||||
3306 sGlnGlnValAlaGlnAlaGlnAsnValAlaGlyValAspThrValLys 3323
|||||
2987 TCAAC...AATACCGGCAAGAACCGCTAGTCTCGACAAATTAAGCGTA 3033
|||||
3323 eSerAlaIleAsnThrLeuAsnGlyAlaMetGlyThrLeuArgAsnSerIle 3339
|||||
3034 GTGGAAGGAAAGACACACACCGCTGTCCGAAATCTTAATTTTC..... 3078
|||||
```

```
3340 .....GlnAspAsnThrAlaThrAsnAsnGlyGlnAsnTyrLeuAs 3353
|||||
3079 .....ACCTGCAAAAGACACACGTCGATGCCGCG 3109
|||||
3353 pAlaThrGluSerAsnLysThrAsnTyrAsnAlaValAspSerAla 3369
|||||
3110 CATGGCGTTATGACGCTTATCCGAAAGACGGGAGTCCGCTGCATATAT 3159
|||||
3370 AsnGlyValIleAsnAlaThrSerAsnProAsnMetAspAlaAsnAlaI 3386
|||||
3160 CCGGTCAAGAAACAAAGAGCTTTCGCAAACTCGCGCAAGCGCGAGAAAC 3209
|||||
3386 e.AsnGlnIleAlaThrGlnValThrSerThrLysAsnAla..... 3399
|||||
3210 AGAGCGCGCTTGACGCGCAAAACAGGCACTTCCCGCAAAACAGG 3259
|||||
3400 .....LeuAspGlyThrHisAsnLeuThr.GlnAlaLysGlnThrA 3413
|||||
3260 CGGAAAAA.....GACACGCGCGCAAGCCTT.....GAC 3288
|||||
3413 LThrAsnAlaIleAspGlyAlaThrAsnLeuAsnLysAlaGlnLysAsp 3429
|||||
3289 GCGCTGATTGCGCGCGCGCAATGCCAACCGAAAGCGAAGATGTTGC 3338
|||||
3430 AlaLeuLysAlaGlnValThrSerAla...GlnArgValAlaAsnValThr 3445
|||||
3339 CGAACCGCGCGCGCACAGCAGCGGGGAAATCCGCGCATTAATGCAGCGG 3388
|||||
3445 rSerIleGlnGlnThrAlaAsnGluLeuAsnThrAlaMetGlyGlnLeuG 3462
|||||
3389 AGGAAGAGAAAAACGGGTGCGAGCGGATTAAGACACCGCTTGCGGAA 3438
|||||
3462 LHisGlyIleAspAspGlnAsnAlaIleThrLysGlnThr.....GlnLys 3476
|||||
3439 CAGCGGAAAGCGAAACCGCGCGCTACACCGCGCTTCCCGCGCGCG 3488
|||||
3477 TyrArgAspAlaGlu...GlnSerLysThrAlaTyrAspGlnAla.V 3492
|||||
3489 CCGCGCGCGCGGATTTCCCGCAACCGGACGCCCAACCGCAACCCGAC 3538
|||||
3492 AlaAlaIleAlaLysAlaIleLeuAsnLysGlnThrGlySerAsnSerAsp 3508
|||||
3539 CCGAGCGGACCTGATACAGCGCTTATGCCAATAGCGGTTTGATGAATTT 3588
|||||
3509 LysAlaAlaValAlaAspArgAlaLeuGlnGlnVal..... 3519
|||||
3589 TCCGCCAGCTCAACAGCGTTTCCCGCTACAGGCAATTTGACCGCGCT 3638
|||||
3520 .....ThrSerThrLysAspA 3525
|||||
3639 GTTTGCGGAAGACCGCGCAACCGCTTGGCAAGCGGATCCGGGACA 3688
|||||
3525 LaleuAsnGlyAspAlaLysLeuAlaGluAlaLysAlaAlaIle..... 3539
|||||
3689 CCAAACTATACCTTCCGCAAGATTTCCGCGCTACCGCAACAAACGAC 3738
|||||
3540 .....LysGlnAsnLeuGlyThrLeuAsnHisIleTh 3550
|||||
3739 CTGCGGCAAAATGGTATGCGAAGAAACCTGGCGACGCGCGCTGCGCAT 3788
|||||
3550 rAsnAla.....GlnArgThrAlaLeuGluGly..... 3559
|||||
3789 CCGTTTTCGCACAACCGGACCGGAAACACTTCGACGACGCGCATCGGCA 3838
|||||
3560 .....GlnIleAsnGlnAlaThr 3565
|||||
3839 ACTCGCACGCGTTGCGCACGCGTGCCTTTCGGGCAATACGCGCATCGCG 3888
|||||
3566 ThrValAspGlyValAsnThrValLysThrAsnAlaAsnThrLeuAspG 3582
|||||
3889 AGGTTTCGACATCGGCATCAGCG.....CGGCGCGCGGTTTATG 3926
|||||
3582 ValaMetAsnSerLeuGlnGlySerIleAsnAspLysAspAlaThrLeuA 3599
|||||
```



640 .....CGCAAGTTCATATCATATTGCAAC..... 666  
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGlu 386  
667 .....GCATATTCTTGG.....CTGCTGGTGCATA 694  
386 sglYglYargAlaIleValItrGlyAspIleAlaLeuIleAspLysnI 403  
695 CCTTGCACAAATGATCAGTGTGGTGCACAGCACTTATGTAAGCAA 744  
403 IeAsnAlaGln...GlySerGly..... 409  
745 AAAATTAAACATAGCCCATATGTTTTCACACAGAGGCTCATTTGG 794  
410 .....AspIleAlaLysThrGlyLysPheValGI 419  
795 CGACAGTGGCTCACCATGTTATCTATGATGCCCAAAAGCAAAAGTGT 844  
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429  
845 TAATTATGGGTATGTCACACAGCAACCCCTATATAGAAAGCAAT 894  
430 .....AsnAlaIleValAspAla..... 435  
895 GGCTTCCAGCTAGTTCGTAAGATGTTCTATGATGAATCTTGGCTGG 944  
436 .....LysGluTrpLeuAspPro..... 442  
945 AGATACCATTCAGTATTCCTACGAACCAATCAAAATGGGAATACCTTT 994  
443 .....AspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerg 459  
995 TTAACGACATATATATGCGCGCAGAGAAATCGATGCCAAACATAAACAC 1044  
459 IuaSpspGluTyrThrGlySerGly.....AsnSer 469  
1045 TATCTCTACTTATAGATTAAACACGACGCTTCAATGTTTAATGT 1094  
470 AlaserThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485  
1095 TTCTTTATCCGAG.....ACAGCAA 1114  
485 rThrLeuGluSerIleLeuLysGlyLysPheValAsnIleThrAla 502  
1115 GAGAACTGTTTAT..... 1128  
502 snGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518  
1129 .....CATGCTGCAGGTGGGTCAACAGTTATGCAAC 1160  
519 ThrLeuTrpSerGluGlyArgSergLysGlyVal..... 530  
1161 CAGACTGAATTAAT.....GGAGAAATA 1183  
531 GlnIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547  
1184 TTCTCTTATTTGACAAAGAAAGTGAATTGATTACTTACACGACATC 1233  
547 eutThrIleTyrSergLysGlyTyrValAsp.....ValHisLysAsnIle 561  
1234 AACCAAGCGCGGCGGCTTTGTATTTGAGGTAATTTTACGGTCTGCC 1283  
562 SerLeuGlyAla.....GlnLysAsnIleAsnIleThrAl 573  
1284 TAA.....AACACGAAAGCTGCCAAGCG 1309  
573 alYsGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590  
1310 CGGCGCTTCATATCATGATGCGAGTACGCTTACTTGAAGTAAGCGC 1359  
590 InGlyThr...IleThrSergLysAsnGlnLysGlyPheArgPheAsnAsn 605  
1360 GTGCAAAACGACCGCTGTCCAAATGCGCAAGGACGCTGTGTTCA 1409  
  
606 ValSer.....LeuAsnGlyThrGlySergLys...LeuGlnPheThr 618  
1410 AGCCAAAGGGGAAAC.....CAAGGCTCGG 1435  
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635  
1436 TCAGCGTGGCGCGAGTAAAGTATCTTA.....GAT 1467  
635 euAsnIle...SergLysValAsnIleSergMetValLeuProLysAsn 650  
1468 CACGACGCGGATCATCAGGCAAAACACGCTT..... 1503  
651 GluSergLysThrAspLysPheLysGlyArgThrTyrTrpAsnLeuThse 667  
1504 .....AGTAAATCGGC.....TTGGTCACGCGCAGG 1531  
667 rLeuAsnValSergLysGlyLysPheAsnLeuThrIleAspSergArg 684  
1532 GAGCG.....GTG 1539  
684 LysSerAspSergAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyLe 700  
1540 CAAGTGAATGCCGATATCATGATCAACCCGACAA..... 1575  
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717  
1576 .....CTCT 1579  
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsn 734  
1580 ATTTCGCTTTCGCGCGCGCGCTTGGATTGACGCGCATTCGCTTTCG 1629  
734 YrLaserPheAsnGlyAsnIleSergValSergLysGlyLysSergValAsp 750  
1630 TTCACCGCAT.....CAAAATACGATGAAGGCGCATGATTGT 1670  
751 PheThrLeuLeuLaserSerSergAsnValGlnThrProGlyValValI 767  
1671 CAACCAAT.....CAACACAAAG 1690  
767 euAsnSerLysTyrPheAsnValSergThrGlySerSerLeuArgPheLys 784  
1691 AATCCACCGTTACCATTCACGCG.....AATAAGATTAATCT 1728  
784 hSergLysThrThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800  
1729 ...ACAACGCGCATATAC..... 1743  
801 AsnAlaThrGlyLysAsnIleThrLeuGlnValGluGlyThrAspGI 817  
1744 .....AACAACTTGATAGCAAAAGAAATTCGCTCAACAGCT 1782  
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGly 834  
1783 .....TGTTTGGCGAGAAAGATGCA...ACCAAAAGCAAGCGCGG 1821  
834 LysAsnIleThrPheGlySergArgLysAlaValThrGluIleGluLysn 850  
1822 CTCATCTGAATTTACCAACCGGAGAGCGGATCGCACTTTA..... 1863  
851 ValThrIleAsn...AsnAsnAlaAsnValThrLeuIleGlySe 864  
1863 ..... 1863  
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIle 881  
1864 .....CTGCTTTCGCGCGCA...ACAAATTTAAACGCG 1893  
881 IeAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly 897  
1894 AATATACG...CAACAAAGCGCAACTGTTTTCAGCGCA..... 1933

898 AsnLeuThrValGluSerAsn..AlaAsn...PheLysAlaIleThrAsn 913  
 1334 .....GAGCGACACCGCAGCCCTACATC 1957  
 913 heThrPheAsnValGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929  
 1958 ATTTAGCAAGCGGTGTCAAAATAATGCAAGGTATCCCAAGAGAAATC 2007  
 930 SerIleAlaLysGlyAlaArgPheLysAspIleAspAsnSerLysAs 946  
 2008 GTGAGCGCAACAGATTGATGACCGC..ACATTAAAGCGGAAACTTCC 2056  
 946 nLeuSerIleThrAsnSerSerSerThrTyrArg..... 958  
 2057 ATATTCAAGCGGCGACAAAGCGGTTCGCCGATGTTGCCAAGTGGAA 2106  
 959 .....ThrIleSerGlyAsnIleThrAsnLysAsn 969  
 2107 GCGCATTCGATTTAAGCAATCAGCC.....CAAGCATTTT 2144  
 970 GlyAspLeuAsnIleThrAsnGlyLysSerAspThrGluMetGlnIleG 986  
 2145 CGGTGTGCACCGCATCAAGCCACACATCTGTACACGTTCCGACTGGA 2194  
 986 yGlyAspValSerGlnLysGlyAsnLeuThrIleSerSerAspLysI 1003  
 2195 CGGGCTGTACAAAGTTGTACGAAATAACCATTTACC...GAGCATTAATG 2241  
 1003 leAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015  
 2242 ATTCCTTCATTGAGCAGACCGCAGACATCAAGCAATGTCAAGCTTGGCGA 2291  
 1016 AspGlyGlnAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrI 1032  
 2292 TCAGGCT.....CATTTAAATCTCAGACTT. 2319  
 1032 eLysThrLysGlnLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhe 1049  
 2320 .....GCCACATCAAGCGC 2344  
 1049 snLysAlaGlnIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065  
 2335 AATCTTACGAGCGGAGACAGCAGCTATACGCTTACGCGCAAGCCGAC 2384  
 1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082  
 2385 CCAAAAGCGCAACCTCAAGCTGTGGCAATGCCAAGCAACATTTATC 2424  
 1082 LlysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099  
 2435 AAGCCACATTAACGCGCAACATCGCTTCGCAATGCTTCATTTAAT 2484  
 1099 yValGlnThrSerGlySerAsnAsnAsnThrGlnHisSerSerAspAsn 1115  
 2485 CTAGCAACAACGCCGTCACAAACGCGAGTGTGACGCTTCCGCAACGC 2534  
 1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnIle 1132  
 2535 TAAGCAAC.....GTAAGCATTCGCGACATCAAGCGCAATGTC... 2574  
 1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyIleThrT 1149  
 2575 .....TCCCTACCGATAAGCAGATTCATTTGAAAACAGC 2613  
 1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGlnIleThr 1165  
 2614 CGCTTACCGGAAATACGCGGCGCAAGATAC.....GCATT 2654  
 1166 AlaGlnThrGlySerIleLeuGlyLysIleGluSerSerSerGlySerVa 1182  
 2655 ACATTAAAGACAGCAGATGAGCGCTCCGTCGGGACAGGAATTAAGCA 2704  
 1182 lThrLeuThrAlaThrGlnGlyAlaLeu.....AlaValSerA 1195

2705 ATTTAAACCTTGACACGCGACCATTTACATCAATTCGCCCTATGACAC 2754  
 1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207  
 2755 GATGCGGACGCGCGCAACCGGACAGTGGCGAGATGCGCGCGCGCGG 2804  
 1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216  
 2805 TTGCGCGCGTCCCTATTTATTCGTTACGCGCGCATTCGCGAGATGCC 2854  
 1217 ThrIleLysGlyThrGlnSerValThr.....ThrSerGlnSerG 1231  
 2855 GTTTCACACGCTGACGCTAACGCGCAATTAACGCTGAGGACA... 2901  
 1231 Ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242  
 2902 ...TTCCGCTTATGTCGGAATCTTCGCTACCGCAGCGCAATTTGA 2948  
 1243 GluValLysAlaThrGlnSerLeuThrThrGlnSerAsnSerLysIle 1259  
 2949 GCTGCGGAAAGTCCGAGGACCTTACACCTTGCTGTC..... 2988  
 1259 s...AlaThrThrGlyGlnAlaAsnValThrSerAlaThrGlyThrIle 1275  
 2989 .....AACATACCGGCAACGACCGGTAAGTCTGACGCA 3024  
 1275 yGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291  
 3025 TTGACGGTA.....GTGGAAGGAAAGACAA 3050  
 1292 LeuThrValGlyLysnGlyAlaGlnIleAsnAlaThrGlnGlyAlaAlaThr 1308  
 3051 CACACCGCTGCCGAAATCTTAATTCACCTCGCAAAAGCAACAGCTGC 3100  
 1308 rLeuThrThrSerSerGlyLysLeuThrThrGlnAlaSerSerHisIle 1325  
 3101 ATGCGCGCGCATGCGCTTATCAAGCTTATCCGAAAGCGCGAGTTCCGC 3150  
 1325 hrSer..AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValA 1341  
 3151 CTCGATATCCGCTCAAGACAGAGCTTCCGCA 3187  
 1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353  
 seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA01845  
 seq\_documentation\_block:  
 ID AAB01845 standard; Protein: 1180 AA.  
 XX AAB01845;  
 AC AAB01845;  
 DT 11-SEP-2000 (first entry)  
 DE Haemophilus influenzae strain 15 mature HmW2A protein, SEQ ID NO:65.  
 XX  
 DE Haemophilus influenzae strain 15 mature HmW2A protein, SEQ ID NO:65.  
 XX  
 DE Mature HmW protein, hmw gene; hmwA1; hmwA2; high molecular weight;  
 KW non-typable Haemophilus influenzae; NTHI; non-encapsulated;  
 KW recombinant production; Escherichia coli; antibacterial; vaccine;  
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
 KW detection; diagnosis.  
 XX  
 OS Haemophilus influenzae strain 15.  
 OS  
 PN W0200020609-A2.  
 PN  
 XX  
 PD 13-APR-2000.  
 PD  
 XX  
 PF 07-OCT-1999; 99W0-CA00938.  
 PF  
 XX  
 PR 07-OCT-1998; 98US-0167568.  
 PR  
 XX  
 PR 08-DEC-1998; 98US-0206942.  
 PR  
 XX

(CONN-) CONNAUGHT LAB LTD.  
 Loomore SM, Yang Y, Klein MH;  
 WPI; 2000-303789/26.  
 N-PSDB; AAA52194.  
 Nucleic acid molecule for producing recombinant high molecular weight  
 proteins of Haemophilus which are used as a vaccine to provide  
 protection against Haemophilus induced diseases in humans -  
 Claim 8; Fig 27A-R; 307pp; English.

The invention relates to the recombinant production of Haemophilus  
 influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 expression construct used to effect recombinant expression comprises a  
 promoter functional in E. coli (e.g., the T7 promoter) operably linked  
 to a modified hmwaBC operon from a non-typable (non-encapsulated) H.  
 influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene  
 clusters termed hmwa1ABC and hmwa2ABC. Each hmwaBC operon comprises hmwa,  
 hmwb and hmwc genes. The hmwa gene encodes the structural HMWA proteins  
 and the hmwb and hmwc genes encode accessory proteins which are  
 responsible for post-translational processing and secretion of the HMWA  
 proteins. The modified hmwaBC operon used in the expression construct of  
 the invention contains an A gene modified such that it encodes only the  
 mature HMWA. The invention also discloses hmwa genes (AAA52175-52198)  
 and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae  
 strains Joyce, K1, K21, LCD2, PMH1, 15 and 12. The nucleic acids and  
 vectors are used for the production of recombinant H. influenzae HMW  
 proteins which can be used as vaccines to mediate a humoral or  
 cell-mediated immune response to provide protection against diseases in  
 humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 non-typable strains of Haemophilus and clone hmw genes from other  
 CC present sequence represents a mature HMWA protein from a non-typable  
 strain of H. influenzae.

SO Sequence 1180 AA;

alignment\_scores:      Quality: 266.00      Length: 1110  
                          Ratio: 0.494            Gaps: 57  
 Percent Similarity: 48.468      Percent Identity: 21.261

alignment\_block:

US-09-303-518D-653 x AAB01845 ..

Align seg 1/1 to: AAB01845 from: 1 to: 1180

```

574 CCGATGCTGTCGATCGAGCAGCGC..... 600
1  |||:|||||:|||||:|||||:|||||:
   ProGluSnValTyrIleAsnAlaGlyAspAlaGlyArgSerAspThrAs 17
601 .....AGACAATATTGGCGGTCTGATGAA.....GACGACCC. 633
   |||:|||||:|||||:|||||:|||||:
17 nLeuGluAsnGluGlyThrGlyThrGlyGlyUserAlaAspThrProl 34
634 .....AATACCGCGGAAGTTCATATCATATTCACAGCGCATATTCCTGG 678
   |||:|||||:|||||:|||||:|||||:
34 ysaArgAsnAsnAsnThrLysThrThrLeuThrAsnSerThrLeuGluLys 50
679 CTCGTCGTCGTCGATACCTTTCACAA..... 705
   |||:|||||:|||||:|||||:|||||:
51 IleLeuAlaArgGlySerPheValAsnIleThrAlaAsnAsnGluIleAr 67
706 .....AATGATCAGGTGGTGGCAGCTCAACTAGTA 739
   |||:|||||:|||||:|||||:|||||:
67 gvaIAsnSerAspIleAsnIleGlyLysSerHisLeuThrLeuTrpS 84

```

```

740 GCCGAAAAATTAACATAGCCCATATGCTTTTATACCAACAGAGGCTCA 789
   |||:|||||:|||||:|||||:|||||:
84 eSerLysAsnLysAsnSer.....GlyValLeuIleAsnGlyAsnIle 98
790 TTGGCGCAGGTGCTCACCAATGTTATCTAT.....GA 824
   |||:|||||:|||||:|||||:|||||:
99 ThrSerThrAlaAsnGlyAsnLeuThrIleTyrSerSerGlyTrpValAs 115
825 TGCCCAAAAGCAAAAGTGTATTAATGAGGCTATG.....C 862
   |||:|||||:|||||:|||||:|||||:
115 PileHisLysAsnIleThrLeuGluSerGlyArgLeuAsnIleThrThrL 132
863 AAACAGCAACCCCTATATAGCAAAAGCAATGCTCCAGCTAGTTCGT 912
   |||:|||||:|||||:|||||:|||||:
132 ysaGluGlyAspValAlaPheGluLysGlyAsnAsnLeuThrIleThrGly 148
913 AAGATGTTGTTATGATGAAATCTTGGTCGGATACCATTCATATAT 962
   |||:|||||:|||||:|||||:|||||:
149 Glu.....GlyThrIleThrAlaGlyAsnAsnGlyPheAr 161
963 CTACGAA..... 969
161 gpheGluAsnValSerLeuAsnGlyThrGlyThrGlyLeuPheAsnL 178
970 .....CCACATCAAAATGGGAATATC.....TTTAAACGMC 1002
178 euserArgProGluLysAsnAsnSerLeuValThrAsnTyrPheAsnGly 194
1003 AATATATATGCGCGCAGGAAATATC.....GATGC 1031
   |||:|||||:|||||:|||||:|||||:
195 ThrLeuAsnIleSerGlySerValAsnIleSerMetIleProAsnAl 211
1032 CAACACTAATACACTATCTCTACCTTATGATTAACACGACACCGTT 1080
   |||:|||||:|||||:|||||:|||||:
211 aThrSerAsnTyrPyrSer.....ArgTyrLysGlyArgThrTyrT 225
1081 .....CAATGTTTAATGTTTCT 1098
225 rpaSnIleThrHisLeuAsnAlaSerGluAspSerAsnPheAsnLeuThr 241
1099 TTATCGAGACAGCAGACAGACCTGTTATCATGCTGCGAGTGGGTCAA 1148
   |||:|||||:|||||:|||||:|||||:
242 IleAspSerSerAlaGluAsp.....GlySerAlaProLeuLeuSe 255
1149 CAGTTATCAGCCAGACCTGATATATGAGAAATATTTCTTTATATGACA 1198
   |||:|||||:|||||:|||||:|||||:
255 rSerTyr.....ThrLeuAsn.....GlyIleSerPheThrThrA 267
1199 AAGGAAAGGTGAATGTATCTTACGACGACATCAACCAAGCGCGGGC 1248
   |||:|||||:|||||:|||||:|||||:
267 sp.....ThrThrPheAsnValAsnLysAsnAla... 276
1249 GGTGTGATATTTAGAGGTAAATTT.....ACGGT 1277
   |||:|||||:|||||:|||||:|||||:
277 .....LysValAsnPheAsnIleLysAlaProIleGlyThrIle 289
1278 CTCGCGCTAAACACAGCAAGACGTGGCAA.....GGCGCGGCGTTCATA 1321
   |||:|||||:|||||:|||||:|||||:
289 eAsnGluThrAsnAsnLeuAsnTyrAlaLeuPheAsnLysAsnIleSerV 306
1322 TCAGTGTATGGCAGTACCGTTACTTGGAAGTAAAGCGC..... 1359
   |||:|||||:|||||:|||||:|||||:
306 alSerGlyGlyLysAsnValThrPheArgLeuAsnAlaSerSerSerV 322
1360 .....GTGGCAACGACGCGCTGTCCAAATGCGCAA 1391
   |||:|||||:|||||:|||||:|||||:
322 GluGlnThrProGlyValIleIleAsnSerLysHisLeuAsnAlaSerL 339
1392 AGGCACGCGTGTGTTCAAGCCAAAGGGGAAACCAAGCGTCGGACAGCG 1441
   |||:|||||:|||||:|||||:|||||:
339 sGlySerSerLeu.....ArgPheGluThrThrGlySerThrLysV 353
1442 TGGCGCAGCGTAAAGTCATCTTAGATCAGAGCGGACGATCAAGGCATA 1491

```

```

353      |||||
353      a1g1y..... 354
1492  AACCAAGCCTTAGTGAATCGGCTTGTCAGCGGAGGAGGAGGTGCA 1541
355  .....Phele 356
1542  ACTGAATGCCGATATACAGTCAACCCGACAAACTCTATTTCGGCTTTC 1591
356  u1leasnasnasp1euthrleuasn..... 364
1592  GCGGCGAGCTTGATTTGAACGGCAATTCGCTTCGTCACCGCAT 1641
365  .....AlaThrGlyGlyAsn1leSerleuengluval 375
1642  CAAATATCCGAT.....GAAGGCGCGATGATTTGTCACCAACA 1679
376  glu1gly1leaspglymet1leglyglu1y..... 385
1680  TCAGACAAAGATCCACCGCTTACCATTTACGCAATTAAGATTACTA 1729
386  .....ValValAlaLysLysAsn1leThrP 394
1730  CAACCGGCATTAACAACTGGATGAGCAAAAA.....GAATT 1770
394  heThrGlyGlyAsn1leThrPheGlySerLysLysAla1leThrGlu1le 410
1771  GCCTACCAACGGTTGGTTGGCGAGAAA...GATCAACCAAAAGAACGG 1817
411  LysGlyLysVal1Thr1leasnGluasnThrAsn1leThrleu1leGlyse 427
1818  GCGGCTCAAT.....CTGAATTACCAACGGAGAAAGCGG 1852
427  TASPPhasnaspHisLysLysProleuAsn1leLysGlyAspValVala 444
1853  ATCGCACTTACTGCTTTCCGCGGAGACA.....ATTTAAACGGCAAT 1896
444  snArGlyAsnleuThrAlaGlyLysVal1leAsn1leGlyGlyAsn 460
1897  ATCCGCAAAACAAACGGCAAA.....CTGTT 1922
461  leuThrVal1GluasnGlyAlaAsnleuLysAla1leThrAsnPhThrP 477
1923  TTTGACGGCGAGACCGACCGCATCAATCATTTAGAAC.... 1968
477  eAsnValGly.....GlyLeuPheAsnAsnLysGlyAsnsera 490
1969  .....GGGTGTCAAAAATGAAAGGTATCCCAAGA 2001
490  sn1leSer1leAlaArg1yGlyAlaLysPheLysAsp1leAsnAsnThr 506
2002  .....GAATCGTGTGGACAGCATTTGATGACCGCACATTAA 2042
507  SerSerleuasn1leThrThrAsnSerAspThrThrTyraThrThr1le1 523
2043  AGCGCAAACTTCATATTCAGGCGGA..... 2070
523  eGluGlyAsn1leThrAsnLysAlaGlyAspLeuAsn1le1leAspAsnL 540
2071  .....CAAGCGGTGTTCCCGCAATGTTGCCAAAGTGAAGGC 2109
540  YsglyAsnAlaGlu1leGlu1leGlyLysAsn1leSerGluLysGluGly 556
2110  GATGGCATTTAAGCATTCACGCCACAGCATTTTC..... 2145
557  AsnleuThr1leSerSerAspLys1leAsn1leThrLysGlu1leThr1 573
2146  .....GGTGTCCACCGCATCAAGCACACATCTGTACAGCTTGG 2188
573  elysLysGlyValasnGlyLysAsnserAspser...SerThrLysSerG 589
2189  ACTGACGCGGTCTGACAAAGTTGTACCGAAAAAACCATTTACGAGATAAA 2238
      : : : : :
      : : : : :

```

---

```

589  InAlaAsnleuThr1leLysThrLysGlu1leuLysleuThrGluAspLeu 605
2239  GTGATTTGCTTATTGACAGACCGACATCAGAGCAATGTACACCTTGC 2288
606  Asn1leSerGlyPheAsnLysAlaLys1leValAlaLysAspSer..... 620
2289  CGATCAGCTCATTTAAATCTACA...GGACTTGCCACACTCAACGCA 2335
621  .....SerAsnleuThr1leGlyAsnSerAspSerGly 633
2336  ATCTTAGTGCAGGCGGAGACACGACATATACGTTAGCGGACCGCAC 2385
633  snThrSerAla.....LysThrVal1ThrPheAsnAsnVal 644
2386  CAAACGGCAACCTCAGCCTGTGGCAATGCCAGACATTTATATCA 2435
645  LysAspSerLys1leSerAlaAspLys...LysVal1ThrLeuAsnSe 660
2436  A.....GCCACATTAACGGCAACACATCGCTTCGACAAATCTTCAT 2479
660  rLysVal1LysThr1leuSerAspAsnAspAsnAsnThrGluGlySerA 677
2480  TTAATCTAAGCACACCGCGTACAAAGCGCAGTCTGACCTTCCGAC 2529
677  spAsnAsnThrGlyLeuThr1leThrAlaLysAspValGluValAsnAsn 693
2530  AACGCTAAGGCAAC.....GTAGCATTCGCGCATCAACGGCAATGT 2573
694  Asn1leThrSerHisLysThrValAsnValSerAlaAlaasnGlyGly1 710
2574  C.....TCCCTAGCCGATTAAGCGCATTTTCATTTTGA 2608
710  erThrLysThrGlyThrThr1leAsnAlaThrAlaGlyAsnValGlu1 727
2609  ACAGCGCGTTTACCGGAAAAATCACGCGCGCAAGATACGCAATTAAC 2658
727  leThrAlaHisThrGlySer1leGluGlyGly.....1leGlu 739
2659  TTAAGAAGACGCAATGACGCGCGCTGCGGACGCA.....TT 2699
740  SerLysProGlySerVal1Thr1leValAlaGlyLysAspThr1leuAla 756
2700  AGCAATTTAAACCTTGACACCGCACCATTTACATCTACCTTCGCTATC 2749
756  1GlyAsn1leSerGlyAsnAlaVal1ThrVal1ThrAlaAsnSer..... 770
2750  GACAGCATCGCGACGCGCAACCGGACGACGCGAGATCGCGCGGC 2799
771  .....GlyAlaLeuThrThr1leAlaGlySer..... 779
2800  CGCGCTTCGCGCGTTCCTTATTCCTTACGCGCGCAACTTGGCAGA 2849
780  .....Thr1leLysGlyThrGluSer1leThr.....ThrSerSerG 792
2850  ATCCGCTTTCAACAGCGCTACGCTAAGCGCAATTTGAACGCTGAGGA 2898
792  nSerGly.....Asn1leGlyLys1leSerGlyLysThrV 805
2899  ..ACATTCGCTTTATGCGAACTTCGCTACCGCAGCGCAAAATG 2946
805  alAsnVal1LysAlaThrAsnSerleuThrThrGluAlaAspSerLys1le 821
2947  AACCTGCGGAAAGT..... 2961
822  GluAla1leThrGluGlyLysAlaAsnVal1ThrSerLysThrSer1le1leG 838
2962  .....TCGAAGGCACTTACACCTTGGCTGTCAACATTAACGGCA 3001
838  yGlyThr1leSerGlyGlyThr.....ValGluVal1ThrAla1 851
3002  ACGAACCGTAAGTCTCGAGCAATTAAGCGTAGTGAAGAGAAAGAACAC 3051
851  hreGluGlyLeuThrThrGluAlaGlySerThr1leThrGlyThrGluSer 867

```



```

3052 ACACCGCTGTCGAA.....ATCTTAATTCACCTTCGAAACGA 3092
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
868 VAlTrHrTnSerSergInSergIysnIleGlyMetIleSergIyl 884
3093 ACACCTGCATGCGCGCGCATGCGCTTATCCGCAAGAGCGCG 3142
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
884 yLysValGluValSerAlaThrLys...AspLeuIleThrIysSerg 899
3143 AGTTCGCGCTGCATATCCGTCAGACAGAGCTTCGCAACATC 3192
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
900 .....SergIleIleValAlaThrAla 906
3193 GGCAAGCGGGGAGAAACAGAGCGCGC..... 3219
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
907 GlyIuValAlaSnValThrSerAlaThrGlyThrIleAspGlyThrIle 923
3220 .....TTGACGGCAAAACAGGACACACTTCGCGCAAC 3253
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
923 rGlyAsnThrValAlaSnValThrAlaAsnThrGlyAspLeuThrValGlu 940
3254 AACAGCGCGAAAGACAGACGCGCAAGCTTGACGCGCTGATTGCGGCGC 3303
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
940 sPAlaIleValIleAspAlaThrGlyAlaAlaThrLeuThrAlaThr 956
3304 GGGCGCAATGCCACCGAAAGCGCAGAAAGT 3333
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
957 SerGlyLysLeuThrThrThrLysAlaSerSer 966

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA01844
seq_documentation_block:
ID   AAB01844 standard; Protein; 1188 AA.
XX
AC   AAB01844;
XX
DT   11-SEP-2000 (first entry)
XX
DE   Haemophilus influenzae strain 15 HMW2A protein, SEQ ID NO:63.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typhable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 15.
XX
PN   M0200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999; 99WO-CA00938.
XX
PR   07-OCT-1998; 98US-0167568.
PR   08-DEC-1998; 98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM, Yang Y, Klein MH;
XX
DR   MPI: 2000-303789/26.
DR   N-PSDB; AAS52193.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 27A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a

```

```

CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typhable (non-encapsulated) H.
CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMWA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMWA. The invention also discloses hmwA genes (AAS2175-A52198)
CC and HMWA proteins (AAB01824-B01849) from the non-typhable H. influenzae
CC strains Joyce, KI, K21, LCD22, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typhable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents an HMWA protein from a non-typhable strain of
CC H. influenzae.
XX
SQ   Sequence 1188 AA;

```

```

alignment_scores:      Quality: 266.00      Length: 1110
                        Ratio: 0.494      Gaps: 57
Percent Similarity: 48.468      Percent Identity: 21.261

```

```

alignment_block:
US-09-303-518D-653 x AAB01844 ..

```

```

Align seg 1/1 to: AAB01844 from: 1 to: 1188

```

```

574  CCTGATCGTGTTCGCAATCGAGCAGCG..... 600
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
9  ProGluAsnValIlyrIleAsnIleGlyAspAlaGlyArgSerAspThrAs 25
601  .....AGACAAATATTGGCGGCTGTGATGAA.....GACGAAACC. 633
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
25  nleuGluAsnGluGluTyrThrGlyThrGlyGluSerAlaAspThrProL 42
634  ....AATAACCGCGGAAAGTTCATATCATATTGCAAGCCCATATTCTTG 678
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
42  ysaArgAsnAsnAsnThrLysThrThrLeuThrAsnSerThrLeuGluLys 58
679  CTCGTCGGTGGCAATACCTTTCACAA..... 705
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
59  IleLeuAlaArgIlySerPheValaIleThrAlaAsnGluIleArg 75
706  .....AATGATCAGGTGGTGGCAGCAGTCAACTATTAGGTA 739
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
75  gValaIleAsnSerAspIleAsnIleGlyLysSerHisLeuThrLeuTPS 92
740  GCGAAATAAATTAACATPAGCCCATATGTTTATTACCAACAGAGAGCTCA 789
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
92  eIseIysAsnLysAsnSer.....GlyValLeuIleAsnGlyLysIle 106
790  TTTGGCAGCAGTGGCTCACCAATGTTTATCTAT.....GA 824
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
107  ThrSerThrAlaAsnGlyLysLeuThrIleIlyrSerSergIlyTrpValAs 123
825  TGCCCAAAAGCAAAAGTGTATTAATGCGGTATTG.....C 862
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
123  pIleHisLysAsnIleThrLeuGluSerLysArgLeuAsnIleThrThrL 140
863  AAACAGCAACCCCTATATGAGAAAGACAGGCTTCAGCTAGTGTGT 912
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
140  yGluGlyAspValaIlePheGlyLysGlyAsnAsnLeuThrIleThrIly 156

```



653	LysAspSerIysIleSerAlaAspGlyHis...	LysValThrLeuAsnSe	668
2436	A.....GCCAATTTAAACGGCACACATCGCTTCGCAAAATGCTCAT		2479
668	ILysValIysThrLeuSerAspAsnAspAsnThrGluGlyGlySerA		685
2480	TTAATCAACCAACACACGGCGTACAAAGGCGACGTCGACGCTTCGAC		2529
685	spsAsnThrGlyLeuThrThrIleThrAlaLysAspValGluValAsnAsn		701
2530	AACGCTAAGCGCAAC.....GTAAGCATATCCGACATCCACAGCAAGT		2573
702	AsnIleThrSerHisIysThrValAsnValSerAlaAlaAsnGlyGlyI		718
2574	C.....TCCCTAAGCCGATTAAGCAGATATTCATTGAA		2608
718	eThrIleThrGlyThrThrIleAsnAlaThrIleGlyAsnValGlu		735
2609	ACAGCGCGCTTACCGGAAAAATCAGCGGCGGACAGATACGCAATTAC		2658
735	IeThrAlaHisThrGlySerIleGlnGlyGly.....IleGlu		747
2659	TTAAAGACAGCAATGACGCGCGCGCGGACGAGAA.....TT		2699
748	SerIysProGlySerValThrIleValAlaGlyGlyAspThrLeuAlaVal		764
2700	AGGCATTTAAACCTTGACACAGCCACCATTAACCTCAATTCGCGCTATC		2749
764	IGlyAsnIleSerGlyAsnAlaValThrValThrAlaAsnSer.....		778
2750	GACACGATGGGAGCGCGGCGCAACCGGACGAGTGGCGGACATGCGCGCGC		2799
779	.....GlyAlaLeuThrThrLeuAlaGlySer.....		787
2800	CGCGTTCGCGCGCTCCCTATTATTCGTTACGGTACGCGCACATCTGGGAGA		2849
788	.....ThrIleGlySerIleThrIleThr.....ThrSerSerG		800
2850	ATCCGCTTTCAACACGCTGACGCGTAAAGCGCAATTTGAACGCTAGAGA		2898
800	uSerGly.....AsnIleGlyGlyLysIleSerGlyThrVal		813
2899	..ACATTCCGCTTATGTGCGAACTCTTCGCGTACCGACGCGCAATTG		2946
813	AlaAsnValLysAlaThrAsnSerIleuThrThrGlnAlaAspSerLysIle		829
2947	AAGCTGGCGGAAGT.....		2961
830	GluAlaThrGluGluGluAlaAsnValThrSerLysThrSerIleIleG		846
2962	.....TCGAGAGCACTTACACCTTGCGCTGCACAATACCGGCA		3001
846	yGlyThrIleSerGlyGlyThr.....ValGluValThrIat		859
3002	AGCAACCCGTAACTCGAGCATATGACGGTATGTGGAAGAAAGCAAC		3051
859	hTrpGluLeuThrThrGlnAlaGlySerThrIleThrGlyThrGlySer		875
3052	ACACCGCTTCCGAA.....AATCTTAATTTACCCGCAAAAGCA		3092
876	ValThrThrSerSerGlnSerGlyAsnIleGlyGlyMetIleSerGlyG		892
3093	ACACGTGATGCGCGCGCATGCGCTTAACCTTAATCGTACGCAAAAGCGCG		3142
892	LysValGluValSerAlaThrLys...AspLeuIleThrLysSerGly		907
3143	AGTTCGCGCTGATATCCGGTCAAAAGACAAAGACTTTCGCAAAATC		3192
908	.....SerGluIleLysAlaThrAla		914
3193	GGCAAGGCGGAGAAACAGAGCGCG.....		3219
915	GlyGluValAsnValThrSerAlaThrThrIleAspGlyThrIleSe		931

```

seq_name: /SISL/gcgdata/geneseq/genesegp-emb1/AAL1993.DAT:AAR41725
seq_documentation_block:
ID   AAR41725 standard; Protein; 1536 AA.
AC   AAR41725;
XX
XX
XX
DT   26-APR-1994 (first entry)
DE
DE   High molecular weight protein 1 [HMW1].
HMW: high molecular weight protein; virus; vaccine; influenza;
epitope; immunity; haemophilus influenzae; gene cluster.
KM
XX
OS   Haemophilus influenzae.
XX
XX
Key      Location/Qualifiers
FH       Misc-difference 668..677
FT       /note= "Possibly incorrect sequence. Alternative
FT       sequence for this region is LNVSESEFN.
FT       (See comments).".
PN       W09J319090-A.
PD       30-SEP-1993.
PP       16-MAR-1993; 93WC-USO2166.
PR       16-MAR-1992; 92GB-0005704.
PA       (BARE/) BARENKAMP S J.
PA       (INRM ) INSRM INST SANTE & RECH MEDICALE.
PI       Barenkamp SJ;
DR       WPI; 1993-320683/40.
DR       N-PSTD; AAC49508.
XX
XX   High molecular weight surface proteins - of non-typeable
XX   haemophilus which exhibit immunogenic properties
XX
XX   Claim 3: Figure 2/10; 100pp; English.
XX
XX   The isolation and purification of the high molecular weight protein
XX   enables the identification of the major protective epitopes of the
XX   protein by conventional epitope mapping. These epitopes can then be
XX   synthesised using standard techniques and incorporated into fully
XX   synthetic or recombinant vaccines. This sequence is claimed to be
XX   the same as that given in AAR41723 (High molecular weight protein 1)
XX   although it does differ slightly. (Repeated regions which are
XX   possibly incorrect and occur in the corresponding nucleotide coding
XX   sequence contribute to these differences).
XX
Sequence     1536 AA:

Alignment_scores:
    Quality: 266.00          Length: 1297
    Ratio: 0.439            Gaps: 73
Percent Similarity: 46.723 Percent Identity: 20.971

```



1531 .....GGGAGC.....GTG 1539  
684 lySeraspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700  
1540 CAATGATGCGCGCATATCACTTAACCCGACAAA..... 1575  
701 SerPheAsnLysAspThrThrPheAsnValGlnArgAsnAlaArgValAs 717  
1576 .....CTCT 1579  
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734  
1580 ATTTGGCCTTCGCGGCGAGCGTTGGATTGAACGGGCGATTCGCTTCG 1629  
734 yrlAspSerPheAsnGlyAsnIleSerValSerGlyGlySerValAsp 750  
1630 TTCACCGCGCAT.....CAAAATACCGATGACGGCGGATGATTGT 1670  
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValAlI 767  
1671 CAACGACAT.....CAAGACAAAG 1690  
767 eAsnSerLysThrPheAsnValSerThrGlySerSerLeuArgPheLysT 784  
1691 AATCCACCGTTACCATTAACAGC.....AATAAGATATTACT... 1728  
784 hrSerGlySerThrLysThrGlyPheSerIleGlyLysAspLeuThrLeu 800  
1729 ...ACAACGGCGCATTAAC..... 1743  
801 AsnAlaThrGlyLysAsnIleThrLeuLeuGlnValGlnGlyThrAspG 817  
1744 .....ACAACCTTGGATGACAAAGAAATGCTTACACCGCTT 1783  
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGlnGlyG 834  
1784 GG.....TTGGCGGACAAAGATCCA...ACCAAAACGAAACGGCGG 1821  
834 LysMetArgPheGlySerArgLysAlaValThrGlnIleGlnGlyAsn 850  
1822 CTCATCTGATTTACCAACGGGAGGATGCGCACTTA..... 1863  
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864  
1863 ..... 1863  
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIle 881  
1864 .....CTGCTTCCGGCGGA.....ACAAATTTAAACGGC 1893  
881 leAsnSerGlyAsnLeuThrAlaGlyLysAsnIleValAlaAsnIleAlaGly 897  
1894 AATATTCAGC...CAACAACGCGCAACCTTTTTCAGCGCA..... 1933  
898 AsnLeuThrValGlnSerAsnAlaAsn...PheLysAlaIleThrAsn 913  
1934 .....GACGACACCGCGACCGCTTACATC 1957  
913 heThrPheAsnValGlyLysLeuPheAspAsnLysGlyAsnSerAsnIle 929  
1958 ATTAGAGAACCGGCTGTCAAAAATGGAAGTATCCACAAGAGAATC 2007  
930 SerIleAlaLysGlyLysAlaArgPheLysAspIleAspAsnSerLysAs 946  
2008 GTGTGGACACAGATTGATGACACCGC.ACATTTAAAGCGGAAACCTTC 2056  
946 nLeuSerIleThrThrAsnSerSerThrTyrArg..... 988  
2057 ATATTACGGCGGACAAAGCGGTGTTCCGCAATGTTGCAAAAGTGAA 2106  
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969  
2107 GCGCATTTGCAATTAAAGCATCAACGC.....CAACAGTTT 2144  
970 GlyAspLeuAsnIleThrAsnGlnGlySerAspThrGlnMetGlnIleG 986  
2145 CGGTGCGACCGCATCAAGCCACACAACTGTACACGCTTCGACCTGA 2194  
986 yclAspValSerGlnLysGlnGlyLysLeuThrIleSerSerAspLysI 1003  
2195 CGGGTGTGACAAAGTTGTACGAAAAACCATTC...GACGATTAATG 2241  
1003 leAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015  
2242 ATTGCTTCATTGACCAAGCCGACATCAAGGCAATGTCAGCTTCGCGA 2291  
1016 AspGlyGlnAsnSerAspSerAspAlaThrAsnAlaAsnLeuThrI 1032  
2292 TCACGCT.....CATTAAATCTCACAGACTT. 2319  
1032 eLysThrLysGlnLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhe 1049  
2320 .....GCCACACTCAACGCG 2334  
1049 snLysAlaGlnIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065  
2335 AATCTTAGTCAGCGCGACACGACGACTATACGTTACCGCGACGCGAC 2384  
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVa 1082  
2385 CCAAAACGGCAACCTCAGCTCGTGCGCAATGCCAAGACATTTAATC 2434  
1082 LysAspSerLysIleSerAlaAspGlyAlaLysValThrLeuHisSerL 1099  
2435 AAGCCCATTTAAACGCAACACATCGCTTCGACAAATGCTTCAATTAT 2484  
1099 yValGlnThrSerGlySerAsnAsnThrGlnAspSerSerAspAsn 1115  
2485 CTAGACACACGCGCTACAAAACGCGACTGACGCTTTCGACACGCG 2534  
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnI 1132  
2535 TAAAGCAAC.....GTAGCATTTCCGCACTCAACGCGCATGTC... 2574  
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGlnIleThrT 1149  
2575 .....TCCCTAGCGCATTAAGCGAGTATTCATTTGAAAAACAGC 2613  
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGlnIleThr 1165  
2614 CGCTTACCGGAAAATCAGCGCGCGCAAGATACG.....GCATT 2654  
1166 AlaGlnThrGlySerIleLeuGlyIleGlnSerSerGlySerVa 1182  
2655 ACACCTAAAGACAGGAGTGAAGCGCTGCCGCGGACGAGATTAAGCA 2704  
1182 ThrLeuThrAlaThrGlnGlyAlaLeu.....AlaValSerA 1195  
2705 ATTTAAACCTTGACAAACGACCATTTACACTCAATTCGCGCTATGACAC 2754  
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207  
2755 GATGCGGCGCGCGCAACCGCGAGTGGCGGACGATGCGCGCGCGCG 2804  
1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216  
2805 TTCGCGCGCTCCCTATTTATCCGTACGCGCGCATTCGCGAGAAATCC 2854  
1217 ThrIleLysGlyThrGlnSerValThr.....ThrSerSerGlnSerG 1231  
2855 GTTCAACACGCTGACGTTAAACGCAATTAATGAAGTACAGGACACA... 2901  
1231 Ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242  
2902 ...TTCCGCTTATGTGGAATCTTTCGCGTACCGCAGCGCAATTTAA 2948

```

1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerIlely 1259
2949 GCTGGCGGAAGTCCGAGGACTTACACCTGGTCTC..... 2988
1259 s...AlaThrThrGlyAlaAsnValThrSerAlaThrIleG 1275
2989 .....ACAAATACCGGCAACGAAACCCGTAAGTCTCGACAA 3024
1275 LylGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291
3025 TTGACGGTA.....CTGCAAGAGAAAGACAA 3050
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThr 1308
3051 CACACCGCTCGGAAATCTTAATTCACCCGCAAAAGCAACGACGTCG 3100
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIle 1325
3101 ATGCCGGCGCATGCGCTTATCCGCAAAAGACGGAGTTCGC 3150
1325 hrser.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341
3151 CTGCATATCCGTCGAAGAACAGACGTTCCGACA 3187
1341 aglySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353
seq_name: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:AAU37120

```

```
seq_documentation_block:
```

```
ID AAU37120 standard; Protein; 2344 AA.
```

```
AC AAU37120;
```

```
DT 14-FEB-2002 (first entry)
```

```
DE Staphylococcus aureus cellular proliferation protein #1290.
```

```
XX Antisense; prokaryotic cellular proliferation protein;
```

```
KM antibiotic; antibacterial; drug design.
```

```
XX Staphylococcus aureus.
```

```
PN WO200170955-A2.
```

```
PD 27-SEP-2001.
```

```
PF 21-MAR-2001; 2001WO-US09180.
```

```
XX 21-MAR-2000; 2000US-191078P.
```

```
PR 23-MAY-2000; 2000US-206848P.
```

```
PR 26-MAY-2000; 2000US-207727P.
```

```
PR 23-OCT-2000; 2000US-242578P.
```

```
PR 27-NOV-2000; 2000US-253625P.
```

```
PR 22-DEC-2000; 2000US-257931P.
```

```
PR 16-FEB-2001; 2001US-269308P.
```

```
XX (ELIT-) ELITRA PHARM INC.
```

```
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
```

```
PI Yamamoto RT, Xu HH;
```

```
XX WPI; 2001-611495/70.
```

```
DR N-PSDB; AAS54979.
```

```
XX New polynucleotides for the identification and development of
```

```
PT antibiotics, comprise sequences of antisense nucleic acids -
```

```
XX Example 3; Seq ID No 12713; 511p; English.
```

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are

Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences).

Sequence 2344 AA:

alignment\_scores:                      Length: 1581  
Quality: 266.00                      Gaps: 63  
Ratio: 0.358  
Percent Similarity: 47.059            Percent Identity: 17.710

alignment\_block:  
US-09-303-518D-653 x AAU37120 ..

Align seg 1/1 to: AAU37120 from: 1 to: 2344

```

163 CGCGCATTTGCCGAATAATTAAGCAAGTTTGCAGTCGGCGCAAGATAT 212
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
20 ArgLeuTyrLysSerGlyLysAsnThrValLysSerGlyIleLysGluI 36
213 TGAGGTTTACAACAAAAAAGGAGTGTGCGCAAAATGATGACGAAG 262
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
36 eGluMetPheLysIleMetGly.....L 44
263 CCCCGATGATGATTTTCTGTGATCGGTACGGC..... 300
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
44 euProheIleSerHisSerMetValSerGlnAspAsnIleSerIleSer 60
301 .....GTGGCGCATTTGGCGGGCATCA 323
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
61 LysLysMetThrGlyTyrGlyLeuLysThrThrAlaValIleGlyAla 77
324 ATATATTGTAGCGGTGCGACATAC..... 348
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
77 aPheThrValAsnMetLeuHisAspGlnGlnAlaPheAlaAspAspA 94
349 .....GGCGGCTAT 357
94 laProLeuThrSerGluLeuAsnThrGlnSerGluThrValGlyAsnGln 110
358 AACAACTTGATTTTGTGCGGAGGAGCAACATCCCGAT..... 396
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
111 AsnSerThrThrIleAspAlaSerThrSerThrAlaAspSerThrSera 127
397 ....CAGCACCGCTTTCTTACCAATT..... 420
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
127 lThrLysAsnSerSerValGlnThrSerAsnSerAspThrValSerS 144
421 ..GTGAAGAATAATTAATTAAGCAGGAGCTAACCGCATCTTATGGC 468
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
144 erGluLysSerGluAsnValThrSerThrThrAsnSer..... 156
469 GCGGATTATCATATGCCGCTTGCACAATTGTTCACAGATGACGAAC 518
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
157 .....ThSerAsnGlnG 161
519 TGTTGATGACCAAGTATATGATGCGTGAATAACGCTGATTAATA 568
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
161 nGluLysLeuThrSerThrSerGluSerThrSerLysAsnThrThS 178

```

569 AATACCTGATCGGTTCGAATGGAGGACGACAAATATTGGCGCT 618  
||| |||  
178 eierSerApPmLysSerValThSer .....ThSer 189  
619 GATGACAGCAGCAACCCATAACCGGAAGTTCATCATATTGCAAGC.. 666  
::: ::::: ||| ::::: |||||  
190 SerThrGluGlnProIleAsnThrSerThrAsnGlnSerThrAlaSerS 206  
666 ..... 666  
206 nasnThrSerGlnSerThrThrProThrSerAlaAsnLeuAsnLysThr 223  
667 ..... GCATATCTCTGG 678  
223 erThrThrSerThrSerThrAlaProValLysLeuArgThrPheSerArg 239  
679 CTCGTCGGTGGCAATACCTTGGCAGAAATGATCA.....GG 716  
||| ::::: ||||| ::::: :::::  
240 Leu...AlaMetSerThrPheAlaSerAlaIleThrThrThrAlaLeuH 255  
717 TGGGCGACAGTCACTTAGTACGCAAAAAATTAACTATGCCCATATG 766  
::: ||| ::::: ||||| ::::: |||||  
255 rAlaAsnThrIleThrValAsnLysAspAsnLeuLysGln..... 268  
767 GTTTTTCACCAAGAGGAGCTATT.....GGCAGCAGTGGCTCA 807  
::: ||| ::::: ||||| ::::: |||||  
269 ..TyMetThrThrSerGlyAsnAlaIleThrTyAspGlnSerThrGlyVal 284  
808 CCAATGTTTATCATGATGCGCCAAAGCAAAAGTGGTAAAT.....<sup>†</sup> 849  
::: ::::: ||||| ::::: |||  
285 ValThrLeuThrGlnSprThrThrSerGlnLysGlyAlaIleThrLeuG1 301  
850 .....AATGGGTATTGGCAA 865  
301 yThrArgIleAspSerAsnLysSerPheHisPheSerGlyValAsnL 318  
866 CAGCAACCCCTTATATGAGAAAAAGACAGCTCCAGCTACTTGCTAA 915  
||||| ||| ||| ::::: |||||  
318 euGlyAsnLysTyTGluGlyHsGlyAsnGly..... 328  
916 GATTGTTCTATGATAAATCTTTGGTGAGATACCATTCAGTATGTA 965  
329 .....GlyAspGlyIleGlyPheAlaPh 336  
966 CGAACCAATCAAAATGGAATACTTTTAAAGCAATAATTAATGGCG 1015  
::: ||| ::::: ||||| ::::: |||  
336 eSerProGlyValIleGlyGluThrGlyLeuAsnGlyAlaIleValGly1 353  
1016 CAGA.....AAATGATGCCCAACATATA 1041  
||| ||| ::::: ||||| ::::: |||  
353 IeGlyGlyLeuSerAsnAlaPheGlyPheLysLeuAspThrHis... 368  
1042 CACTATTCTCTACTTATAGATTAAACACAGACCGTTCATTTGTTAA 1091  
369 .....AsnThrSerThrProAsn..... 374  
1092 TGTTCTTTATCCGAGACAGAGAACCTGTTATCATGCTGCAGTGG 1141  
||| ::::: ||||| ::::: |||||  
375 ....SerSerAlaLysAlaLysAlaAspProSerAsnValAlaGlyG 390  
1142 GGGTCACACGTTATGACCCACACGTAATATGAGAAAAATATTTCCTT 1191  
||| ::::: ||||| ::::: |||||  
390 LysAla.....PheGlyAlaPheValThrThr 398  
1192 ATTGACAAAGAAAGGTGAATTGATCTACTACCAACATCAACCAAG 1241  
::: ||| ::::: ||||| ::::: |||||  
399 AspSerTyGlyValAlaIleSerThrTyThrSerSerSerThrAlaAsp 415  
1242 CGGCGCGGTTGTATTTTGAAGGTAAATTTACGCTGCTGCCTAAACAA 1291  
||||| ||| ::::: ||||| ::::: |||||  
415 naIleAlaLysLeu.....AsnValGlnProThrAsnA 426  
1292 ACGAACACGTGGCAGGCGGCGCTCATATCAGTCAGTCAGTACCGCTT 1341

[illegible]

```

641 .....ValValSerThrAspGlnAla 647
2167 CACACACATCTGACACGCTTCGGACTGACAGGGCTCGACACAGTTGTACCGA 2216
      ::::: ||| ::::: ::::: ::::: |||
648 AsnAsnLysSerThrThrThrPheThrIleAsnValValAspThrThrAl 664
2217 AAAAACCATTATAC .....GAGATATAAGTGAATGCTTCAT 2251
      ||::: ||| ::::: ::::: |||
664 aProthrValThrProIleGlyAspLysSerSerGlnValPheSerProI 681
2252 TGACCAAGACCGACATCAGAGGCAATGTCACGCTTGGCCATCAGCCTCAT 2301
      ::::: ||| ::::: |||
681 leSerProIle.....AsnIleAlaThrGlnAspAsnSerGly 693
2302 TTAAATCTCAAGAGACTTCGCCACACTCAACAGGCAATCTTAATGAGGGCG 2351
      ::::: ||| ::::: |||
694 AsnAlaValAlaThrAsnThrValThr .....GlyLeuProSerGly.. 706
2352 AGACACGACCTATACGGTTACGGTCAGCGACGACCGACCCAAAC.....G 2392
707 .....LeuThrPheAspSerThrAsnAsnThrIleSerG 718
2393 GCAACCTCAGCCTCGGGGACATGGCCAAAGAACATTAATCAAGCCGCA 2442
718 LThrProThrAsnIleGlyThrSerThrIleThrIleValSerThrAsp 734
2443 TTAAACGGCAAC.....AC 2456
735 AlaSerGlyAsnLysThrThrThrThrPheLysTryGluValThrArgAs 751
2457 ATGGCGTTGGGACAAATGCTTCATTTAATCTAAGCAACACCGCCGACAAA 2506
751 nSerMetSerAspSerValSerThrSerGlySerThrGlnGlnSerGln 768
2507 ACGGCAGTCGTGACGCTTCCGACACACGCTAAGAGCAACGTAAGCATTCC 2556
768 eAlSerThrSerLysAlaAspSerGlnSerAlaSerThrSerThrSer 784
2557 GCACCTCAACGGCAGATCTCTCCTAGCCGATTAAGCAGATTCATTGTA 2606
785 .....GlySerIleMetThrSerThrSerAla..... 793
2607 AAACACCGCGTTACCGGAAATAACGAGCGGCGGACAGATACGACATTAC 2656
794 .....SerThrSerLysSerThrSerValS 802
2657 ACTTAAAGACACGCGATGAGACGCTGCCG.....TCGGACACGGAATTA 2700
802 eLysSerAspSerValSerAlaSerLysSerLeuSerThrSerGlnSer 818
2701 GGGCAATTTAACTTCACAAAGCAGCATTACACATTCATTCGGCTATCG 2750
819 AsnSerValSerSerSerThrSerThrSerLeuValAsnSer..... 832
2751 ACACGATGCGGCGACGGCGCAAAACCGGACAGTGGCGAGATGCG..... 2793
833 GlnSerValSerSerMetSerGlySerValSerLysSerThrSerL 849
2794 .....CCGGCCCGCGCTTCGGCGCGGCTTCCTCATTAATCCGTT 2829
849 euSerAspPheIleSerAsnSerSerSerThrGlnLysSerGlnSerVal 865
2830 ACGCCCGCACTTCGGGACAAATCCGTTTCAACAG..... 2865
      ::::: ||||| ::::: ||| ::::: |||
866 SerThrSerThrSerAspSerLeuAlaGlnThrSerThrSerLeuSerAsp 882
2866 .....CTGACGGTAAACGGCAAAATTGACGCGTACAGGACATTCCGCT 2908
882 ValSerMetSerThrSerGlySerLeuSer..... 892
2909 TTATATGCGCAACTCTTCGGCTACCGAGCGGCAAAATTGAACCTGGCGGAA 2956
893 .....LysSerGlnSerLeuSerThrSerThr 901

```

```

2959 AGTTCCGAAGGACACTTACACCTTGCCCTGTCAAACTATCCGACGACGAC 3000
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
902 SerAspSerIleSerThrSerGlnSerValSerIleSerThrSerAsnSe 918
3009 CGTAAGTCTCGAGCAATTAGCGGTAGTGGAGGAGAAAGACAAACACACCGC 3058
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
918 ILeSerThrSerGlnSerIleSerGlnSerGly.....SerThrSerg 933
3059 TGTCGCAAAATCTTAATTTCACCCCTCGCAAAAC..... 3090
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
933 ILeSerIleSerIleSerIleSerIleSerIleSerValSerIleSerThr 949
3091 GAACACGTGTGATCCGCGCGACGATGCGCTTATACGCTTATCCGCAAAAGCG 3140
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
950 SerIlySLeuGlnSerGlnSerThrSerIleSerIleSerThrSerAspSe 966
3141 CGAGTTCCGCGCTGCATTAATCCGGGTCAAAAGAACAAAGACTTTCGACAAAC 3190
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
966 IlyS.....SerMetSerThrSergIleSerIleSerThrSerAspSerT 979
3191 TCGCGCAAGCGCGGAGAAACAGAGCGCGCTTGACGCGCAAAACAGACGACAA 3240
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
979 hSerThrSerIleSerValSerGlySerIleSerValIleGlySergIle 995
3241 CTTTCGCGCAACACACAGCGCGGAAAAACACACCGCGCAAGACTT..... 3285
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
996 SerValSerThrSerThrSerAspMetSerThrSergIleMetIleSe 1012
3286 .GACGGCGCGATTTGGGCGCGGCGGCAATGCCACGCAAAAGCAGCAAAAGTG 3334
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1012 rAspSerMetSerThrSergIleSerLeuAlaIleSerIleSerIleSerm 1029
3335 TTGCGCCAAACGCGCGCGGACGAGCGCGGGAATAATGCCGCGATTATGACG 3384
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1029 eSerValSerSerSerMetSerThrSergIleSergIlySerThrSergIle 1045
3385 GCGGAGAGAAAGAAAAACGGGTGCAGCGCGATTAAGACAC..... 3426
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1046 SerLeuSerAspSerIleSerThr...SerAspSerAspSerIlySergIle 1061
3427 .GCCTTGGGAAAAACAGCGGAGCGGAAACCGCGCGCTTACACACGCGCT 3475
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1061 uSerIleSerThrSergIleSergIlySerThrSerThrSerThrSerThrS 1078
3476 TCCCGCGCGCGCGCGCGCGCGCGGATTTCGCGCAACCGCACGCCCA 3525
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1078 eSerSerSerValIly..... 1082
3526 CCGCAACCCCAACCGGACGAGCGACCTGATCAACCGCTTATGCCAATAGCG 3575
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1082 ..... 1082
3576 TTTCAGTGAATTTTCGCGACGCTCAACAGCGTTTTCGCCCTACAGACG 3625
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1083 .MetSergIleSergIleSerThr.SergIlySerMetSerThrSergIleSer 1098
3626 AATTGGACCGCGTGTTCGCGAAAGCGCGGACCGCGCTTTGGACGACG 3675
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1099 AspSerThr.....SerIleSerThrIleSerPheSerAspSe 1110
3676 GGCATCCGGGACACCAACACTACGCTTGCGCAAGATTTCCGCGCGCTACG 3725
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1110 rThrSerAspSerIly.....SerIleSerThrA 1120
3726 CCAACAACACCGACTCGCCCAATGGGTATGCAGAAAAAACTCGGACAGC 3775
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1120 lAserSergIleSerIleSergIleSerValSerThrSerThrSergIlySer 1136
3776 GGGCGCGTGGCAACGCTGTTTGGCAACACCGGACCGGAAACACCTTCGAC 3825
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1137 ValSerThrSerThrSerIleSerIleSerThrSerAsnSergIleuIlyThrSerTh 1153

```







```

2299 CATTAAATTCACAGGACTTGGCACAACGTCAGCGCAATCTTAATGACAG 2348
      ||| ..... ||| ..... ||| ..... |||
620 Aspleu... IleIleGlyAsnSerAspAsnAsnAsnAlaAsnAla... 634
2349 CGGAGACAGCGACTATACGGTTACGCGCACAGCGCCAAACAGCGCAACC 2398
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
635 ..... LysLysValThrPheAsnGlnValLysAspSerLysI 647
2399 TCACCTTCGCGGCAATGCCAAGCAACATTATATCA...GCCACA 2442
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
647 IeSer...AlaAspSerHisAsnValThrLeuAsnSerLysValGluThr 662
2443 TTAACGCGACACATCGGCTGGACAAAT.....GC 2474
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
663 SerAsnLysAsnAsnAspAlaGluSerAsnAsnGlyAspLysThrSerLe 679
2475 TTCATTATATCTAACAACAACGCGCTCAAAACGCG..... 2511
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
679 uThrIleAsnAlaLysAsnIleThrValAsnAsnIleThrSerHisL 696
2512 .....ACTGACGCTTTCGACAC.....GCTAAGCGCAACGTA 2547
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
696 yStrValAsnIleThrAlaSerGlnAsnValThrThrLysAlaGlyThr 712
2548 AGCCATTCGCGACTCAACGCGCAATGTCCTGACCGATTAAGGACGATAT 2597
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
713 ThrIleAsnAlaThrThrGlySerValGluValThrAlaLys..... 726
2598 CCATTGTAACACAGCGCTTACCGGAAATACGCGGCGGCAAGATA 2647
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
727 .....ThrGlyAspIleLysGlyLysValGlu 736
2648 CGGATTAACACTTAAGACAGGGAATGACGCTGCCGTGGCGACAGGA 2697
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
736 eThr.....SerGlySerValThrLeuThrAlaThrGlyL 748
2698 .....TTAGCAATTTAAACCTTGACACAGCCACATTAACACTCA 2738
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
749 AlaLeuAlaValSerAsnIleSerGlyAsnThrValThrIleThrAlaAs 765
2739 TTCGCGCTATCGACAGCATCGCGACAGCGCGCAACCGCGACGTGGCAG 2788
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
765 nLysGlyLysLeuThrThrGlnAlaGlySer...ThrValSerAlaIleA 781
2789 ATGCGCGCGCGCGCGCTTCCGCGCTCCCTATTATCCGTAGCGCGCCA 2838
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
781 snGlyValThrAlaSerSerGlnSerGlyAspIleSerGlyThr..... 795
2839 ACTTCGGACAAATCCGTTCAACACGCGTACGCGTAACGCGCAATTTGAA 2888
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
796 .....IleSerGlyAsnThrValLysValSerAlaI 806
2889 CGGTACGAGGACATTCGCTTTATGTCGGAACCTTCGCTACCGACGC 2938
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
806 eGlyAspLeuThrThrLysSerGlySerGluIle..... 817
2939 GCAAAATGAGCTGCGGAA.....AGTTCGCAAGCGCACTTAC 2976
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
818 ..LysAlaLysThrGlyGluAlaAsnValThrSerAlaThrGlyThrIle 833
2977 .....ACCTTG.....GCTGTCAC.....AATACGCGCAA 3002
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
834 GlyGlyThrIleSerGlyAsnAlaValAlaThrAlaAsnThrGlyAs 850
3003 CGAACCGGTA.....AGTCTCGAGC 3022
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
850 pLeuThrValGluAspAlaAlaLysIleAspAlaThrGlyGlyAlaAla 867
3023 AATTCAGCGTATGAGAGAAAGACACACACCGCTGTCGCAAAATCTT 3072
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
867 hTrLeuThrAlaThrSerGlyLysLeuThrThrLysAlaSerSerSerIle 883

```

```

3073 AATTACCCCTGCAAACAGACACGTCGATGCCGCGCATGGCGTTATCA 3122
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
884 Thr.Ser.....AlaAsnAsnGlnValA 891
3123 GCTTATCCGCAAGACGCGAGTTCGCTGATATCCGTCACAAAGAAC 3172
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
891 snLeuSerAlaLysAspGlySerIleGlyGlyAsnIleAsnAlaAlaAsn 907
3173 AAGAGCTTTCGCAACAACCTCGCA 3196
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
908 ValThrLeuAsnThrThrGlyAla 915
seq_name: /SIDS1/gcdata/geneseq/geneseq-emb1/AA2000.DAT:AA01834
seq_documentation_block:
ID AAB01834 standard; Protein: 1101 AA.
XX
AC AAB01834;
XX
DT 11-SEP-2000 (first entry)
DE Haemophilus influenzae strain LCDC2 HMW1A protein, SEQ ID NO:43.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typhable Haemophilus influenzae; NHI; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain LCDC2.
XX
PN W0200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99MO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
XX
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PS Loosmore SM, Yang Y, Klein MH;
XX
WP1: 2000-303789/26.
XX
N-PSDB: AAA52183.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 12; Fig 22A-P, 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typhable (non-encapsulated) H.
CC influenzae (NHI). Most HMW-expressing NHI strains contain two hmw gene
CC clusters termed hmwA1ABC and hmwA2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMW
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMW. The invention also discloses hmwA genes (AAA52175-AA52198)
CC and HMW proteins (AAB01824-B01849) from the non-typhable H. influenzae
CC strains Joyce, KI, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as

```

CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMW protein from a non-typeable strain of  
 CC H. Influenzae.

xx Sequence 1101 AA;

alignment\_scores:  
 Quality: 265.50 Length: 1042  
 Ratio: 0.503 Gaps: 55  
 Percent Similarity: 50.672 Percent Identity: 21.881

alignment\_block:  
 US-09-303-518D-653 x AAB01834 ..

Align seq 1/1 to: AAB01834 from: 1 to: 1101

```

574 CCTGATCGTGTTCGATCGGAGCAGACAAATATTGGCGTGTGATGA 623
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7  ProSpluValThrIleGlyAlaGly..AspValGlyArgSerSps 22
624 AGAC.....GAACCCAAATAACCGCG 643
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 pSerSerAspThrAlaPheProThrGlyThrGlyGluArgAsnSerPro 39
644 AAGTTATATCATATTGCAAGCGCATATCTTGGCTGCGTGGGCAAT 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 yStrAsnAlaGlnAsnArgProThrIleThr.....Asn 50
694 ACCTTGCACAAATGATGATGAGTGGTGACAA...GTCAACTTAGGTAG 740
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 ThrSerLeuGluGlnIleLeuLysAsnGlyThrPheValAsnIleThrAl 67
741 CGAA.....AAATTAACATAGCCCAT 763
|||||:|||||:|||||:|||||:|||||:|||||:
67 alyAsnLysIleLeuValAsnSerAspIleAsnIleLysGluAsnSerH 84
764 AT.....GGTTTTTATCCACA 780
|||||:|||||:|||||:|||||:|||||:|||||:
84 IsteuIleLeuThrSerGluArgSpsGlyAsnSerGlyValGlnIleAsp 100
781 GGAGGCTCATTTGGCAGACAGTGGCTCACCAATGTTATCTAT..... 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 GlyAsnIleThrSerAlaThrGlyGlySerLeuThrValTyrSerSerG 117
823 .....GAGGCCAAAGCAAAAGTGTAAATTAATGAGGTATTG.... 861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 yTrpValAspValHisLysAsnIleThrLeuAsnSerGlyTyrLeuAsn 134
862 .....CAACAGGCAACCCCATATATAGAAAGCAATGGCTTCAG 903
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 IeThrThrLysSerGlyAspValAlaPheGluGlnGlyAsnAspLeuThr 150
904 CTAGTT.....CGTAAGATTGGTTCATGA 929
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 IleThrGlyGlnGlyThrIleThrAlaSerLysLysGlyPheArgPheAs 167
930 TGAATCTTGGCTGGAGATACCATTCATGATTTC.....TAGCAAC 970
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 pAsnValThrLeuSerGlyValLysLysGlyPheLeuPheLysTyrSerG 184
971 CACATCAAAATGGGAAA.....TACTTTTAAACGACAAAT 1005
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 IuThrAsnAsnLysAspSerAsnPheGluAsnHisPheArgGlyThr 200
1006 AATATGGCGCAGAAATATGAT.....GCCAAACATAAACA 1043
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 LeuAsnIleSerGlyLysValAspIleLeuMetGlnAlaArgGlnGluAs 217
1044 CTATTCTCTACCTTAT.....AGATTAA 1066

```

```

217 nTrpAsnArgArgHisSerGlyArgSerHisIleTrpAsnValThrArgLeuA 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1067 AAGACAGCAACGTCGATGTTATGTTCTTTCTTCAGACAGCAAGCA 1116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 nValSerThrAsnSerTyrLeuAsnIleThrIleAspAsnSerGlySer 250
1117 GAACSTGTTTATCATGTCGACAGTGGGGTCAACAGTTATGACCCGACT 1166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 ArgPro.....SerProGlyAlaGlyProLeuTyrArgArgSerG 264
1167 GAATATGCA.....GAAATATTTCCTTTATGACAAAGGAAAG 1207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 yLeuAsnGlyIleSerPheAsnAsnAspThrValAlaPheAsnValAlaSerG 281
1208 GTGAATGATA.....CTTACCGCAACATCAAC 1236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 ySerAlaValAsnPheSerIleLysProProIleValSerAlaValHis 297
1237 CAAGCGCGCGCGGCTTGTATTGAGGGTAAAT..... 1269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 ArgGlyAsnHisThrLeu...PheAsnGlyAsnValSerValLeuGly 313
1270 .....TTTACGCTCTCGCTTAAAAACAAACGAAAGT 1300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 yGlyAspValAsnPheHisPheAsnAlaSerSerSerHisIleStrThrH 330
1301 GG.....CAAGCGCGCGCGCTTCATATGATGATGATGACGATGAC 1338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
330 IseGlyValValIleLysSerGlnAsnPheAsnAlaSerGluLysSer 346
1339 GTTACTTGGAAAGTAACCGCGTGGCAACGCGCTGTCCAAAATGCG 1388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
347 LeuArgPheLysSerGluGly.....SerThrArgThrAlaPheThrI 361
1389 CAAGGCAACGCTGCTGTTCAAGCCCAAGGGGAAACCAAGCTGTGTC 1438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 eGluSerAspLeuThrLeuAsnAlaThrGly..... 371
1439 GCGTGGCGCGGTAAGTCAATCTATGATGACGACGCGGACGATCAAGC 1488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 .....GlyAsnIleSerLeuAsnGlnValAlaGlyIleAspGly 384
1489 AAAAAACAACGCTTATGTAAGTGGCTGTGTCACGCGCAGG..... 1530
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
385 AsnLeuGlnLys.....SerLeuValAlaAsnLysAsnIleThr 397
1531 .....GGACGCGTCACTGATGCCGATATCATGATTCACACCCG 1570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
397 rPheGluGlyGlyAsnIleThrLeuAlaAla.....A 408
1571 ACAAACTATTTTCGCGCTTCGCGCGGACGCTTGGATTGAAACGGCAT 1620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 sPlysLysProIleGluIleLysGlyAsnIleThrValLysGluLysAla 424
1621 TCGCTTTCGTCACCGCATTCAA...AATACGATGAGAGGCGCATGAT 1667
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 AsnValThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeu 441
1668 TGTCAACCAATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441 rIleArgGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerA 458
1717 .....AAGATATTACTACACCGCAAT..... 1740
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
458 IalLeuAsnIleGluLysAsnLeuThrValGluGlySerAlaLysPheLeu 474
1741 .....ACAAACACTT 1751
475 AlaAsnProAsnTyrSerPheAsnValSerGlyLeuPheAspAsnGln 491
1752 GGATAGCAAAAAAAGAAATGCTTACACGCT...TGGTTTGGCGGAAG 1798
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

491	yltysSerAsnIleSerIleAlaIleLysIleGlyIleGlyIleAlaHisPheLysAspIleA	508
1799	ATGCACCAACAAAGCAACGGCGCGCTTAATCTGAATTATCCACCGCAAGCA	1848
508	snasnrThlLysSer.....LeuAsnIleThrThrsnSerAspSer	521
1849	GCAGCATCGCACTTACTGCTTCGCGCGCAACAATTAAAGCGCAATAT	1898
522	AlaIleThrIleAlaIleIleIleGlyAsnIleThrAsnSerAsnGlyAspE	538
1899	CACGCAACAAACGCAACCTGTTTTCAGCGGACAGCCGACCGCAACG	1948
538	uasnIleThrAspAsnLys.....AsnAsnA	547
1949	CTACATCAATCTTTTGAAGACGGGTGCTCAAAATGAAGGATTCCTCCACAA	1998
547	IaGulIleGlnIleGlyIleGlyAsnIleSerGlnLysGlnLys.....	560
1999	GGAGAAATCGTGTGGACACAGATTGGATGCACCCACATTTAAAGCGCA	2048
561	.....AsnLeuThrIleSerSerAs	567
2049	AAACTTCATATTCACGGGGGCAACAGCGTGGTTCCCGCAATGTTGCCA	2098
567	plvslIleAsnIleThr.....AsnGlnIleThrIleLysLysGlyValAsnL	583
2099	AACTGGACAGCGATTTGCGATTTAAGACATACAGCCCAAGCATGTTTCGT	2148
583	ylsGlnspSerAspSerSerThrIleAsnAsnAlaAsnLeu.....	596
2149	GTCGACACCGCATCAACGACCACTCTGTACAGCTTGCGACTGACGGG	2198
597	.....ThrIleLysThrLys.....	601
2199	TCGACCAAGTGTACCGCAAAAACCATTAACGACCATTAAGATGATGCTT	2248
602	.....GlnLeuGlnLeuThrIleLysPheAsnAsnIleSerG	613
2249	CATTGACGAAGACCGCAATCAAGGCATGTACGCTTGCCGATCAGCT	2298
613	lyPheAsnplvslIaGulLys.....ThrAlaLysGlnLysIa	625
2299	CATTTAATCTCAACAGACTTCCCAACATCAAGGCATCTTAGCGAGG	2348
626	AspLeu.....IleIleGlyAsnSerAspAsnAsnAsnAlaAsnAla..	640
2349	CGAGACACGCACTATACGCTTAGCGGACAGCGCACCCAAAGCGCAAC	2398
641	.....LysLysValThrPheAsnGlnValLysAspSerLysI	653
2399	TCAGCTCTGTGGCATGCGCCAAAGCAATTTAAATCAAA.....GCCACA	2442
653	Ileer.....AlaAspSerHisAsnValThrLeuAsnSerLysValGlnLut	668
2443	TTAAACGCAACACATCGGCTTCGAGCAAT.....GC	2474
669	SerAsnGlyAsnAsnAspIaGlnSerAsnAsnGlyAspGlyThrSerIe	685
2475	TTTCATTATCTTAACGACCAACGCGGTCAAAAGGC.....	2511
685	uthrIleAsnAlaLysAsnIleThrValAsnAsnAsnIleThrSerHisL	702
2512	.....AGTCTGACGCTTCCGCAAC.....GCTAAAGCAACGTA	2547
702	ylsThrValAsnIleThrIleAsnSerGlnAsnValThrThrLysAlaGlyThr	718
2548	AGCGATTCGCGCACTCAAGCGCAATGCTTCCTAGCCGTAAGCAGATTT	2597
719	ThrIleAsnAlaThrThylGlySerValGlnValThrAlaLys.....	732
2598	CCATTTTGAAACACGCGGTTTACCGGAAATAACGCGCGGCAAGATA	2647
733	.....ThrGlyAspIleLysGlyLysValGln	742

[illegible]

PF 16-MAR-1993; 93WO-US02166.  
 XX  
 PR 16-MAR-1992; 92GB-0005704.  
 XX  
 PA (BARE/) BARENKAMP S J.  
 XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI  
 XX Barenkamp SJ;  
 DR WPI: 1993-320683/40.  
 DR N-PSDB: AA049506.  
 XX  
 PT High molecular weight surface proteins - of non-typeable  
 XX haemophilus which exhibit immunogenic properties  
 PS Claim 3; Figure 2; 100pp: English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 XX  
 XX Sequence 1536 AA;

alignment\_scores:  
 Quality: 265.00 Length: 1297  
 Ratio: 0.438 Gaps: 73  
 Percent Similarity: 46.646 Percent Identity: 21.126

alignment\_block:  
 US-09-303-518D-653 x AAR41723 ..

Align seg 1/1 to: AAR41723 from: 1 to: 1536

```

178 AATAAGGCAAGTTTGCAGTCGGCGCAAGATATTGAGTTTACACAA 227
    ||| ||| :|||:||||| ||||| :|||
199 ASnHsIleuIlethrhIvalGly...LysaspIlyserVal... 211
228 AAAAGGGAGTTGGTGGCAATATGATGACGAAGCCCCGATGATGATT 277
    |||:|||||:||||| :|||
212 .....AsnIleuIleGlyIlyValIlyAsnGlu..... 221
278 TTCTGTGATATGCGGTAAACGGC.....GTGGCGGCAATTGGCGGGCAT 321
    :|||:||||| ||||| :|||:|||||
222 ..GlyValIleSerValAsnGlyIlySerIleSerIleuAlaGlyGln 237
322 CAA..... 335
238 LysIlethrhIleSeraspIleIleAsnProthrhIlethrhIlySerIleAl 254
336 CGTGGCAGATTAACGCGGCTATTAACAATGTTGATTTGTGCGGAGGAA 385
    :|||:||||| :||| :|||:|||||
254 alaIleProGluAsnGluAlaValAsnIleuGlyAspIlePheAlaIlyGly 271
386 GCAAT..... 411
    :||| :|||:||||| :|||:|||||
271 LysAsnIleAsnValAlaGluAlaIlethrhIleArgAsnGlnGlyIlySer 287
412 TACCAAAATTTGAAAGAAATTAATTATAAGCAGGACTAAGCCCATCC 461
    :|||:||||| :|||:|||||
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGGCGCGCATTTATCATATGCCGCTTGGCCAAATTTGTGCACAGATG 511
    :|||:||||| :|||:|||||
298 ..... 303
512 CAGAACCTGTTGAG.....ATGACCAAGTTATATGAT 543
    :|||:||||| :|||:|||||
303 IatysGluGlyIlyAlaGluIleGlyIlyAlaIleSerAlaGluAsnGln 319
544 GGGTGGAAATACGCTGATTTAAATAAATACCCATGCTGTT.....CG 587

```

```

320 GlnAlaIlyGlyIlyLysLeuMetIlethrhGlyAspLysValthrhLeuLys 336
    ||| :||| :|||:|||||
588 AATCGGAGCA.....GGCAGACAAATATT 610
    :|||:||||| :|||:|||||
336 sthrgIlyAlaValIleAspLeuSerGlyIlyGluGlyIlythrhTyrL 353
    :|||:||||| :|||:|||||
611 GCGGCTCTGATGAA...GACGAACCCAAATAC..... 639
    :|||:||||| :|||:|||||
353 euGlyIlyAspGluArgGlyIlyAsnGlyIleGluIleuAlaLys 369
    :|||:||||| :|||:|||||
640 .....CGCGAAAGTTCATATCATATTCGAAC..... 666
    :|||:||||| :|||:|||||
370 LysThrSerLeuGluIlySerIlethrhIleAsnValSerGlyIlyGlu 386
    :|||:||||| :|||:|||||
667 .....GCATATCTTGG.....CTCGCGGTGCGCATTA 694
    :|||:||||| :|||:|||||
386 sglIyIlyArgAlaIleValIlyPGLyAspIleAlaIleuIleAspIlyAsnI 403
    :|||:||||| :|||:|||||
695 CTTTGCACAAATAGATCAGGTGGTGGCACAGTCACCTTAGTAGCGAA 744
    :|||:||||| :|||:|||||
403 IeAsnAlaGln...GlySerGly..... 409
745 AAATTAACATACCCATATGCTTTTACCACAGAGGCTCATTTGG 794
    :|||:||||| :|||:|||||
410 .....AspIleAlaLysThrGlyIlyPheValG 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGAGTGT 844
    :|||:||||| :|||:|||||
419 uhrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGTATTGTAACACAGCACCCCTATATAGAAAAAACAAT 894
    :|||:||||| :|||:|||||
430 .....AsnAlaIleValAspAla..... 435
895 GCGTTCACACTAGTTCGTAAGATGTTGTTCTATGATGAATCTTGTGG 944
    :|||:||||| :|||:|||||
436 .....LysGluThrPheLeuAsp.....Phe 442
945 AGATACCATTCAGTATCTTACAGAACACATCAAAATGGAATATCTTT 994
    :|||:||||| :|||:|||||
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459
995 TTAACGACAAATATTAATGCGCGCAGAAAAATGATGCCCAACATTAACAC 1044
    :|||:||||| :|||:|||||
459 IAspAspGluThrIlyThrGlySerGly.....AsnSer 469
1045 TATCTCTACCTTAATGATTAATAACACAGACCGTTCAATGTTATATGT 1094
    :|||:||||| :|||:|||||
470 AlaSerThrProLysArgAsnLysGluIlyThrThr..LeuThrAsnTh 485
1095 TTCTTTATCGAG.....ACAGCAA 1114
    :|||:||||| :|||:|||||
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIlethrhAla 502
1115 GAAACCTGTTAT..... 1128
    :|||:||||| :|||:|||||
502 snGlnArgIleThrValAsnSerSerIleAsnLeuSerAsnIlySerLeu 518
1129 .....CATGCTGAGGTGGGTCACACAGTTATACGCC 1160
    :|||:||||| :|||:|||||
519 ThrLeuThrSerGluIlyArgSerGlyIlyGlyVal..... 530
1161 CAGACTGAATAT.....GGAGAAATA 1183
    :|||:||||| :|||:|||||
531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTTATTAACAAGAAAGTGAATGATGACTTCCACAGCAACATC 1233
    :|||:||||| :|||:|||||
547 euthrhIleThrSerGlyIlyThrPvalAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGGTGTTGATTTTGGAGGTAATTTACGCTCGCC 1283
    :|||:||||| :|||:|||||

```

```

562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAA.....ACACAGCAACGTGGCAAGCG 1309
573 alysglnaspIleAlaPheGlnIulysGlySerAsnGlnValIleThrGlyG 590
1310 CGGCGCTTCATATCAGTGTAGCAGTACCGCTTACTTGGAAAGTAAACGCG 1359
590 lngIyThr...IleThrSerGlyAsnGlnIulysGlyPheThrPheAsn 605
1360 GTGGCAACAGACCGCGCTGCCAAATCGCAAGCAGCAGCTGCTGTCA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheThr 618
1410 ACCCAAGAGGGAAC.....CAAGCTCGG 1435
618 rthrIysaThrAsnIulysTyAlaIleThrAsnIulysPheGlnGlyThrL 635
1436 TCAGCGTGGCGAGGTAAAGTCATCTTA.....GAT 1467
635 euAsnIle...SerGlyysValAsnIleSerMetValIleuProIysAsn 650
1468 CAGCAGCGGAGCATCAAGGCAAAACAGACCTT..... 1503
651 GluSerGlyTyAspLysPheLysGlyIarGlyTyTyrPasnIleuThrSe 667
1504 .....AGTAATCGGC.....TTGTCACGCGGAGG 1531
667 rIeuAsnValSerGluSerGlyIulPheAsnIleuThrIleAspSerHArg 684
1533 GCAGC..... 1539
684 lySerAspSerAlaGlyThrIleuThrGlnProTyAsnIleuAsnIlyle 700
1540 CAACGTATCGGATATCAGTTCACCCGCAAA..... 1575
701 SerPheAsnIulysAspThrThrPheAsnValGlnArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnIulysTrSerSerIeuAsnT 734
1580 ATTTGGCGTTTCGCGCGGAGCGGATTTGATTTGAACGGGCGCTTCG 1629
734 yrlaSerPheAsnIulysAsnIleSerValSerGlyGlySerValasp 750
1630 TTCCACCGCAT.....CAAAATCGCATGAAGGCGGATGATGT 1670
751 PheThrIleuAlaSerSerSerAsnValGlnThrProGlyValIle 767
1671 CAACCAAT.....CAAGACAAG 1690
767 eAsnSerIulysTrPheAsnValSerThrGlySerSerIeuArgPheIysT 784
1691 AATCACCCTTACCATTAACAGC.....AATAAGATATTACT... 1728
784 hrSerGlySerThrIulysThrGlyPheSerIleGlnIulysAspIleuThrIeu 800
1729 ..ACAACCGCAATAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuGlnValGlnIulysThrAspG 817
1744 .....ACAACTTGATACAAAGAAAGAAATTCCTACAAAGCT. 1782
817 yMetIleGlyIulysGlyIleValAlaIulysAsnIleThrPheGlnIulysG 834
1783 .....TGCTTGGCAGAGAGATGCA...ACCAAAACGACAGGCGG 1821
834 IysAsnIleThrPheGlySerArgIulysAlaValThrGlnIleGlnIulysAsn 850
1822 CTCATATCGATATCAACGAGAGAGGAGGATGCACCTTA..... 1863
851 ValThrIleAsn.....AsnAsnAlaAsnValThrIleuIleGlySe 864

```

---

```

1863 ..... 1863
864 rAspPheAsnAsnHisGlnIulysProIeuThrIleLysLysAspValIle 881
1864 .....CTGCTTCGGGGGA.....ACAAATTAAACGCG 1893
881 leAsnSerGlyAsnIleuThrAlaGlyIulysAsnIleValAsnIleAlaGly 897
1894 AATATCAG...CAACAAACGCGCAACCTGTTTTTCAGCGCA..... 1933
898 AsnIleuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsn 913
1934 .....GACGACACCGCAGCCTACATC 1957
913 heThrPheAsnValGlyIleuPheAsnAsnIulysAsnSerAsnIle 929
1958 ATTTAGCAAGCGGTGCTCAAAATGAGAGTATCCCAAGAGGAGAAATC 2007
930 SerIleAlaIulysGlyAlaArgPheLysAspIleAspAsnSerIulysAs 946
2008 CTGTGGACAACGATTCGACCGC.ACATTTAAACGGAACCTCC 2056
946 nIeuSerIleThrAsnSerSerThrIulysArg..... 958
2057 ATATTCAAGGCGGACAGCGGTGCTCCCGCAATGTCGCAAGTGAA 2106
959 .....ThrIleSerGlyAsnIleThrAsnIulysAsn 969
2107 GCGCATTCGATTTAAACATCAAGC.....CAACAGCTTT 2144
970 GlyAspIleuAsnIleThrAsnGlnIulysSerAspThrGluMetIleG 986
2145 CGGTGTCACACCGCATCAAGCAACACATCTGACAGCTTCAGCTGCA 2194
986 yGlyAspValSerGlnIulysGlnIulysIleSerSerAspIulys 1003
2195 CGGCTCGACAGTGTACCGAAACCACTTCC...GACGATTAAGT 2241
1003 leAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTCATTCAGCAAGACGACATCAGAGCAATGTCAGCTTCGCA 2291
1016 AspGlyIulysAsnSerAspSerAlaThrAsnAlaAsnIleuThrI 1032
2292 TCACGCT.....CATTTAAATCTCAGAGACTT. 2319
1032 elysThrIulysGluIleuLysIleuThrGlnAspIleuAsnIleSerGlyPheA 1049
2320 .....GCCACACTCAAGCGC 2334
1049 snIulysAlaGluIleThrAlaLysAspIulysAspIleuThrIleGlyAsn 1065
2335 AATCTTAGTCAGCGGAGACGACACTATACGTTACCGGCAAGCCGAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysValThrPheAsnGlnVal 1082
2385 CCAAAACGCGACCTCAGCTGCGGCAATGCCAAGCAACATTTAATC 2434
1082 IulysAspSerIulysIleSerAlaAspGlyHisIulysValThrIleuHisSerL 1099
2435 AAGCCACATTTAAAGGCAACACATCGCTTCGACAAATGCTCATTTAAT 2484
1099 yValGlnIulysThrSerGlySerAsnAsnThrGlnIulysSerSerAspAsn 1115
2485 CTAGCAACAGCGCTGTACAAACGCGACTGTGACGCTTCGCAAGAGCC 2534
1116 AsnAlaGlyIleuThrIleAspIulysAsnValThrValAsnAsnI 1132
2535 TAAAGCAAC.....GTAGCCATTCGCGACTCAACGCAATGTC... 2574
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyIulysIleuThr 1149

```

```

2575  . . . . . TCCCTAGCCGATTAAGCACTATTTCATTGTAAGACAGC 2613
1149  hrlvsthrghlythThlileasnAlathrthrglysnvalGluilethr 1165
2614  CGCTTACCGGAAAATACGCGCGGCAAGATACG. . . . . GCATT 2654
1166  AlaglnthrghlyserlileuglgyglylegluserSerserghlyserVa 1182
2655  ACACCTTAAGACAGCGCAATGACGCGCGCTGCGGCGACGGAATTAGGA 2704
1182  lThleuthrAlathrGluGluAlaLeu. . . . . AlaValSerA 1195
2705  ATTTAAACCTTGAAACGCCACCATCTACTACTATCCGCTATGCACAC 2754
1195  snlIeseglyasnThValThValThraIaasnSer. . . . . 1207
2755  GATCGCGGAGCGCGCAACCGGCGAGTCCGCGAGATGCGCGCGCGCG 2804
1208  . . . . . GlAlaLeuthrThrLeuAlaGlySer. . . . . 1216
2805  TTCGCGCGCTTCCTATTATTCCTATACGCGCGCAACTGCGGCAATTCC 2854
1217  .ThrllelysglythrGluSerValIthr. . . . . ThSerSerGluSero 1231
2855  GTTTCACACGCTGACGCGTAAAGCGCAATTGAACGCTGACGAGACA. . 2901
1231  ly. . . . . AsplleelylythrIleserghly. . . . . GlyThrVal 1242
2902  . . . . . TTCGCTTTATGTGGGAACCTCTGCGGTACCGACGCGCAATTGA 2948
1243  GluValIalathrThrgluSerleuthrThrGluSerasnSerlyIlely 1259
2949  GCTGCGGCAAGTTCGCCAAGCACTTACCTTGCGTCTC. . . . . 2988
1259  s. . . . . AlathrThrglyGluAlaasnValIthrSerAlathrghlyThrIleg 1275
2989  . . . . . AACAACTACGCGCAACCGCGTAAGCTGCGAGCAA 3024
1275  lyGlyThrlIeserghlyasnThrValaasnValIthraIaasnAlaGlyasp 1291
3025  TTGACGCTA. . . . . GTGGAGGAAAAAGACA 3050
1292  LeuthrValIaGlyasnGlyAlaGluIleasnAlathrghGluGlyAlaIath 1308
3051  CACACCGCTGTCCGAAAATCTTAATTTCACCCCTGCAAAAGACACGCTCG 3100
1308  rleuthrThrSerSerghlyLysleuthrThGluAlaSerSerthIster 1325
3101  ATCCGCGGCGCATGCGTATCAGCTATTCGCCAAAAGCGGCGAGTCCGC 3150
1325  hrSer. . . . . AlAlaLysglyGluValaasnLeuSerAlaIaInspgIyservAlaI 1341
3151  CTGCATTAATCCGCTCAAGAACAGAGCTTCCGACA 3187
1341  aGlySerIleasnAlaIaasnValThrLeuasnThr 1353

seq_name: /SIDS1/gcgdata/geneseq/geneseqp_emb1/AA2001.DAT:AAU34143
seq_documentation_block:
ID  AAU34143 standard; Protein; 2086 AA.
XX  AAU34143;
XX  AC
XX  DT
XX  14-FEB-2002 (first entry)
XX  DE
XX  Staphylococcus aureus cellular proliferation protein #419.
XX  KM
XX  Antisense: prokaryotic cellular proliferation protein;
XX  antibiotic; antibacterial; drug design.
XX  OS
XX  Staphylococcus aureus.
XX  GN
WO200170955-A2.

```

```

XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001MO-US09180.
PR
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR
XX N-PSDB; AASS2002.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 5639; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 2086 AA:
SQ
XX
XX alignment_scores:
XX Quality: 261.00 Length: 1529
XX Ratio: 0.345 Gaps: 80
XX Percent Similarity: 49.444 Percent Identity: 22.041
XX
XX alignment_block:
XX US-09-303-518D-653 x AAU34143 ..
XX
XX Align seg 1/1 to: AAU34143 from: 1 to: 2086
XX
XX 289 TCGCCTAACGGCGCTGGCGCATTTGGCGGCAT.....CAATA 326
XX :|||||:|||||:|||||
XX 5 ThrAsnAsnGlyIleThrValThrIleGlyThrPheAsnProAlaAspTh 21
XX
XX 327 TATTTGAGCGTGGACACATACGCGGCGCTATTAACAATGTTGATTTGGTG 376
XX ||| ||||| :|||
XX 21 rIleGlnValValAlaIleThrGlnGly.....Serg 31
XX
XX 377 CGAGGGAAGCAATCCCGATCAGCACGCCGTTTCTTACCAAAATGTG... 423
XX :||| :|||:|||||:
XX 31 IysIuThrIleSerAspGluGlnIleGlySerAspAspPheThrValValAla 47
XX
XX 424 .....AAAAGAATATATTATTAACAGGAGCTAACGGCCAT..... 459
XX :||| ||| |||||

```



```

48  ProGlnProAaSnHsIaSnPrIuSILEtPrGlnAaSnGlyHsIleAa 64
460  ....CCTTATGCGCGGCAATTATGATATG..... 483
      |||  ::  ::  |||:::
64  pIlethPrOaSnAaPrOserGlyHsIleuIleAaSnProHlGlnIleAa 81
      :::::
484  ....CCGGCTTGGCAAAATTTGTCAAGATGCAAGAACCTGTGTGG 525
      :::::
81  etAspIleAaLyThrGlnUysMeGlyAaSnGlyAaGlnHsIleSerUys 97
      :::::
526  ATAGCAGATTATGGATGGG.....TGAATTAATGCGATGT 563
      ::  ::  |||  |||
98  ThrIleAaSnValAaIArgelGylAaSnAaSnGlnPrThIleAa.... 112
      :::::
564  AAATAAATACCCCTGATCGTGTTCGAATCGGACA....GCGACAAT 607
      |||||  |||||  |||  ::  |||
113  .AsnUys...ProAspTyValThrIleuAaSprHlnHsThrGlyUysValT 128
      :::::
608  ATTGGCGGCTGTATGAAGACAAACCAATAAC..... 639
      :::::
128  hrPheAaSnIaAsnThrValUysProAaSnSerAlaIleThrIleThrPro 144
      :::::
640  CGGGAAGTCAATATATATATGAGAGCGGATATTTGCTGGCTGGTGGG 689
      :::::
145  UysAlaGlyThrGlyHsIleSerIaSer..... 154
      :::::
690  CAATACCTTTGCAAAATAGATGACAGTGGTGGGCAAGCTTATAGTA 739
      ||  ::  :::::  |||||
155  .AsnProSerThrIleThrAlaIProAlaIleHrHsThrValAaSnThr..T 170
      :::::
740  GCGAAAAATTAAACATAGCCATATGCTTTTATACCAAGAGGCTCA 789
      :::::
170  hrcIuIleValUys....AspTyGlySerAaSnValThrAlaIaGln 184
      :::::
790  TTGGGCGAGTGGCGTCACCAATGTTTATCAATATGATGCCCAAAACAA 839
      :::::
185  IleAaSnAaIa.....ValGlnValIleAaSnUysArgTh 196
      :::::
840  GTGGCTTAAT...ATGGGGATTTGCAAAACAGCAACCCCTATATAGAA 886
      |||  |||||:::  |||  |||
196  rAlaThrIleUysAaSnGlyThAlaMePrOthrHsIleuAaIaGlyUys 213
      :::::
887  AAAGCAATGGCTTC....CAGCTAAGTTCGTAAGATGTGTTCTATGAT 930
      :::::
213  eArThrHrThrIleProValIThrValIThrUysAaSnArgUysThrGln 229
      :::::
931  GAAATGTTTGGTGAAGATGCGCATACAGATCTTACGAACCAACATCAA 978
      :::::
230  GlUVal.....GlnIuSerIlePheThrUysAlaAspUysArg 242
      :::::
979  ....AAATGGAATACTTTTATACCAACAATATATATGCGCGAG 1018
      :::::
242  gcuIleuIleThrAlaUysAaSnHsIleuAaSprPrOAlaIeSerThrArg 259
      :::::
1019  GAADA.....ATCGATGCCAAACATTAACAC 1044
      ||||
259  IlyUysUysProGlyThrIleThrGlnTyUysAaSnAlaIleHsIaSnIa 275
      :::::
1045  TATTCGTCGCTTATATAGATTAAACACAGAACCGCTTCAATGTATAGT 1094
      :::::
276  GlnGlnGlnIleAaSnThrAlaUysThrIleuAaGlnGlnValIleAaSn 292
      :::::
1095  TCTCTTATAC.....GAGACAGCAAGAGAACCTGTATTC 1129
      :::::
292  ncuIaArAlaThrProGlnGlnValSerAaSprAlaIeUhrIlyUysAlaArg 309
      :::::
1130  ATGCTGCAAGTGGGCTCAACAGTTTGCACCCAGACCTGAATATGAGAA 1179
      |||  :::::
309  lAlaGlnThrUysIleAaSnGlnUysAlaIaIeUeGlnAaSnUysGln 325
      :::::
1180  AATATTCCTTATATGCAAGGAAGATGATGATTAATTCACAGAA 1229
      |||  :::::
326  AspAaSnGlnIleUysValThrSerUysAaSn....LeuGlnIleSer 340
      :::::

```

1230 CACCACCAA...GGCGCGGGCGTTGATTTTGAAGGT.....1267  
340 TAlAsnGlnValProSerThrThreGlyMetHrcGlnGlnSerIleAspA 357  
1268 ATTTTCAGGCTCTCGCTTAATAAACCAACGAAGCTGGCAAGCGCGGCGCTT 1317  
357 snTyrAsnAlaLysLysArgGluIaGluThrGlnIleThrAlaIaGln 373  
1318 CATATCAGTGAATGGCAGTACCGCTACTTGCGAAA.....1350  
374 ArgValIleAspAsnLysAspAlaThrAlaGlnGlnIleSerAspGluAs 390  
1351 .GTAAAGCGCGTGGCAAGACCGCGCTGCGCAAAATCGCAAGACGCAC 1399  
390 nThrThrAlaGlnArgAsnAspLeuThrAsnGlnIleSerGlnAlaThrA 407  
1400 TCGTGGTCAGACCAAGGGAAGCAAGAGCTCGTCAGCGTGGCGCAG 1449  
407 snLeuIaIaValaGluSerValLysGlnSerAlaAsnSerLeu...Asp 422  
1450 GGTAAAGTC.....ATCTAAGATCAG.....1470  
423 GlyAlaMetGlyAsnLeuGlnThrAlaIleAsnAspLysSerGlyThrLe 439  
1471 .....CAGCGGACGATCAAGCAAAACAAACAGCCT 1501  
439 uAlaSerGlnAsnPhleuAspAlaAspGluGln...LysArgAsnAlaT 455  
1502 TTATGTAATTCGGCTTGGTCACGGCGAGGGGACGCGTCAACTGAATGCC 1551  
455 TyrAsnGln.....AlaValSerAsnAlaGluThrIle...LeuAsnLys 468  
1552 GAAATACATCTCAACCCGCAACACTGATTTCCGCTTCCGGCGGACG 1601  
469 GlnThrGlyProAsnThrAlaLys.....ThrAlaValaGln 480  
1602 TTTCGATTTGAACGGGCAATTGCTTTCGCTTCACCGCATTCAAATACCG 1651  
480 uGlnAlaLeuAsnAsnValAsnSerAlaLysIleAlaLeuAsnGlyHrc 497  
1652 ATGAAGGGCGCATGATTGTGCACCAACCAATCAAGCAAAAGATTCACCGTT 1701  
497 In.....AsnLeuAsnAsnAlaLysGlnAlaIaIle 507  
1702 ACC...ATTACAGCAATTAAGATATTACTCAACCGGCAATTAACACAA 1748  
508 ThrAlaIleAsnGlyAlaSerAspLeu.....AsnGlnLysGln 520  
1749 CTTCGATTCACAAAAGAAATTCGC.....TACACG 1780  
520 nLysAspAlaLeuLysAlaGlnAlaAsnGlyAlaGlnArgValSerAsnA 537  
1781 GTTCGTTTGGCGCAAGACGATGCACCAAGCAAGACGGCGCTCAATCTG 1830  
537 IaGlnAspValaGlnArgAsnAlaThrGluLeuAsnThrAlaMetGly... 552  
1831 AATTACCAACGGGAAGACGGAGTCGACTTTCCTTCGGCGCAAC 1880  
553 ThrLeuLysSHSLaIleAlaAspLysThrAsnThrLeuAlaSerSerly 569  
1881 AATTTAAACGCATTCATCAGCAA.....A 1906  
569 snTyrAlaAsnAlaAspSerThrLysGlnAsnAlaTyrThrThrLysValT 586  
1907 CAACGCGCAAACTGTTTTCAGCGGCAAGCCGACA.....1941  
586 hrAsnAlaGlnHisIleIleSerGlyThrProThrValaThrThrPro 602  
1942 .....CGACGCGTCAATCATATTGGAGCGGGGTCGCAAAAT 1982  
603 SerGluValThrAlaIaIaAsnGlnValaAsnSerAlaLysGlnGluLe 619



```

1198 njstYrYrAspAlaIaG...GlnSerLysLysThrLalYrIspGlnA 1214
3485 CCCGCGCGCGCCCGCGGATTTGCCGCAACGCGACGCCCAACGCAACC 3534
1214 Ia.VaIaIaIaIaIaLysAlaIleLeuAsnLysGlnThrGlySerAsnSe 1230
3535 CAACGCGACGCGGACCTCATGCGCGGTATGCGCAATGCGGTTGAGTGA 3584
1230 rAspLysAlaIaIaValAspAlaGlaIaLeuGlnVal..... 1242
3585 ATTTTCCGCGACGCTCAACAGCGTTTCGCCGTACAGACCAATTGGACC 3634
1243 .....ThSerThrLys 1246
3635 GCGTGTTCGCGAAGACCGCGCGCAACGCCGTTTGAGCAAGCGGATCCGG 3684
1247 AspAlaIaLeuAsnGlyAspAlaLysLeuAlaGlnAlaIaLysAlaIaIa.. 1262
3685 GACACCAAAACACTACCTGCTCCAGAGATTTCGCGCCCTACGCGCAACAAC 3734
1263 .....LysGlnAsnLeuGlyYThrLeuAsnHisT 1272
3735 CGACCTGCGCCAAATCGGTATGCAAGAAAACCTCGCAGCGCGCGCTCG 3784
1272 IeThrAsnAla.....GlnArgThrAlaLeuGlnGly..... 1282
3785 GCATCGCTGTTTCGCAACAGCGACCGGAAACACCTTCGACGACGCGATC 3834
1283 .....GlnIleAsnGlnAl 1287
3835 GGCACCTCGCGACGCGCTTCCACGCGTCGCTTCGCGCAATACGCGAT 3884
1287 aThrThrValAspGlyValAsnThrValLysThrAsnAlaAsnThrLeuA 1304
3885 CGGCAGGTTCACATCGGCATCACC.....CGGCGCGCGGTT 3922
1304 SpElyAlaIaMetAsnSerLeuGlnGlySerIleAsnAspLysAspAlaThr 1320
3923 TTAGTAGGCGGACGCTTTCACAGCGCTACAGGACGCAAAA...TCGCGCGC 3969
1321 LeuArgAsnGlnAsnYrLeuAspAlaAspLysSerLysArgAsnAlaYr 1337
3970 CGCGTGTGCAATACGGCATTCACG.....C 3995
1337 rThrIaIaValThrAlaIaIaGlnGlyIleLeuAsnLysGlnThrGly 1354
3996 AAGATACCGCGAGGTTTCGCGGATTCGCGATCGCAACCG...ACATCG 4042
1354 LysAsnThrSerLysAlaIaAspAlaIaAspAlaIaLeuAsnThrValThrArg 1370
4043 GCGCAACGCGCTATTTCGTCGCAAAAAGCGATTACGATACGAAACGTC 4092
1371 AlaLysAlaIaIaLeuAsnGlyAlaGlnAsnLeuArgAsnThrLysThrSe 1387
4093 AATATCGCGACCGCGGCGCTTCGATTACACCGCTACCGCGCGGCGCATTA 4142
1387 r.....AlaThrAsnThrIleAsnGlyLeuP 1396
4143 GGCAGATTATTCATTCAAACCGCGCGCAAC 4171
1396 roAsnLeuThrGlnLeuGlnLysAspAsn 1405

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU37018
seq_documentation_block:
ID AAU37018 standard; Protein; 3158 AA.
XX AAU37018;
XX AC
XX DT 14-FEB-2002 (first entry).
XX DE Staphylococcus aureus cellular proliferation protein #1188.

```

```

KW Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HR;
XX
DR WPI: 2001-611495/70.
XX
DR N-PSDB; AASS4877.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 12611; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3158 AA:
XX
alignment_scores:
XX Quality: 261.00 Length: 1107
XX Ratio: 0.466 Gaps: 51
XX Percent Similarity: 50.587 Percent Identity: 21.951
XX
alignment_block:
XX US-09-303-518D-653 x AAU37018 ..
XX
Align seg 1/1 to: AAU37018 from: 1 to: 3158
XX
1144 GTCAACAGTTATCGACCACGACTGAATAATGAGAAAATTTCCTTAT 1193
XX |||||::: ::: |||||:::|||||:::
XX 818 ValAsnaIaaIatYsgInAlaIeaUsnGIYAasnApsnIleaAlaSnAl 834
XX
1194 T...GACAAGAAGAAAGGTGAATTGCATCTTACACGACATCATCAACCAAG 1240
XX ::::: ||| ::||| :::::::::::
XX 834 alyScInleInProlySgInglInIleaAlaSnLeUrThnHIsIleaAnASP. 850

```



```

1394 ysaSpGlnProGlnThrGlnAlaSerSerLysPheIleAsnGlnAspGln 1410
2875 AACGGCAATTGAACGGTCAGGGACATTCGGCTTATGTCGGAACCTT 2924
      |||
1411 AlAGlnLysAspAlaTyrThrGlnAlaValGlnHisAlaLysAspLeu 1427
2925 CGGCTACCGGACGGCAATTTGAAGCTGGCGGAAGTTCCGAAGCACTT 2974
      |||
1427 easn...LysThrThrAspProThrLeuValLysSerValIleAspGln 1443
2975 ACACCTTGCTGTACACATATCCGGACGACACCCGTAAGTCTCGAGCAA 3024
      |||
1443 LathrGlnAlaValAsnAspAlaLysAsnLeuHisGlnLysAspGlnLys 1459
3025 TTGACGCTAGTGGAGGAAAGACACACACCGCTGTCGGAATCTTAA 3074
      |||
1460 Leu.....AlaGlnAspLysGlnArgAlaThrGlnThrLeuAs 1472
3075 TTTCACCTGCAAAACGACACGTCGATGCCGGCGCATGCGCTTATCAGC 3124
      |||
1472 n...AsnLysSerAsnLeuAsnThrProGln..... 1481
3125 TTATCCCAAGACGCGGATTCGCCCTGCATATCCGTCGAAGAACA 3174
      |||
1482 ....ArgGlnAlaLeuGlnAsnGlnIleAsnAsnAlaAlaThrArgGly 1496
3175 GACCTTCCGCAAACTCGGCAAGCGGAGGAAACAGAGCGCGCTTGAC 3224
      |||
1497 GlnValAlaGlnLysLeuThrGlnAlaGlnAlaLeuAsnGlnIleMetG 1513
3225 GGCAGAACAGGACACACTTGCCTCCGCAACACAGAGCGGAA..... 3264
      |||
1513 uAlaLeuArgAsnSerIleGlnAspGlnGlnThrGlnAlaGlySerL 1530
3265 .....AAAGACAGCGCAAGCCTTGAGCGCGCTGATGTCGCGCC 3303
1530 yspHeIleAsnGlnAspLysProGlnLys...AspAlaTyrGlnAlaAla 1545
3304 GGGCGCAATGCACCGCAAGAGGAGCAAAAGTGTCCGACACCGCGCGCA 3353
      |||
1546 ValGlnHisAlaLysAspLeuIleAsnGlnThrSerAsnProThrLeuAs 1562
3354 GGCAGCGCGGAGAAATCCGCAATTATGACGCGGAGGAAAGAAAAAAC 3403
      |||
1562 pLysAla.....GlnValGlnGlnLeuThrGln 1572
3404 GGGTCGACGGGGATAAAGACACCGCTTGGCGAAACGCGCGCAAGCGGAA 3453
      |||
1572 LysValAsnGlnAlaLysAspAsnLeuHisLysAspGlnLysLeuAlaAsp 1588
3454 ACCCGCGCGCTACACCGCTTCCCGCGCGCGCGCGCGCGCGGGA 3503
      |||
1589 AspLys.....GlnHisAlaValThrAs 1596
3504 TTTCGCCCAACCGCAGCCCAACCGCAACCGCAACCGCGCGGACCTGA 3553
      |||
1596 pLeuAsnGlnLeuAsnSerLeuAsnAsnProGlnArgGlnAla...LeuG 1612
3554 TCAGCGCTTATGCCAATAGCGGTTGATGATTTTCCGCGACCGCTAAC 3603
      |||
1612 LysSerGlnIleAsnAsnAla..... 1618
3604 AGCGTTTCCGCTACAGAGATTCGACCGCGCTGTTTCCGCA..... 3648
      |||
1619 .....AlaThrArgAspGlnValAlaGlnLysLeuAlaGlnAlaG 1632
3649 .....GACCGCGCAACGCGTTTGGACACAGCGGATCCGGGACACCA 3691
      |||
1632 nAlaLeuAspGlnAlaMetGlnAlaLeuArgAsnSerIleGlnAspGln 1649
3692 AACACTACCGT.....TCGCAAGATTTCCGCGCTACCGCGCAACCAAC 3735
      |||
1649 LngLntHnGlnSerGlySerLysPheIleAsnGlnAspLysProGlnLys 1665

```

```

3736 GACCTGCGCAAAATCGTATCGAC.....AAAACCTCGCGACGCGCGC 3779
      |||
1666 AspAlaTyrGlnAlaAlaValGlnHisAlaLysAspLeu..... 1678
3780 CGTCGGCATCTGTTTTCGACACACCGGACCGGAACACTTCG..... 3823
      |||
1679 .....IleAsnGlnThrGlyAsn.ProThrLeuAsp 1688
3824 .....ACGAGCGCATCGGCAAC 3840
1689 LysSerGlnValGlnGlnLeuThrGlnAlaValThrThrAlaLysAspAs 1705
3841 TCGGACGCGC..... 3850
1705 nLeuHisGlnLysAspGlnLysLeuAlaArgAspGlnGlnGlnAlaValThr 1722
3851 .....TTGCCACAGCTGCGCTTTCGGGCAATACGCGATCGGCA 3889
      |||
1722 hValAsnAlaLeuProAsnLeuAsnHisAlaGlnGlnAlaLeuThr 1738
3890 GGTTCGACATCGGCATACGCGGCGGCGGCTTATGACGCGACGCTT 3939
      |||
1739 AspAlaIleAsnAla...AlaProThrArgThrGlnValAlaGlnHisVa 1754
3940 TCAGACGCGCATCAGAGCAAAATCCGCGCGCGCTGTCATACGCGCAT 3989
      |||
1754 LglnThrAlaThrGlnLeuAspHisAla.....MetGlnThrL 1767
3990 TCAGCGACATACCGCGCAGCTTTCGGCGGATTCGCGATCG...AACCGC 4036
      |||
1767 eLysAsnLysValAspGlnValAsnThrAspLysAlaGlnProAsnTyr 1783
4037 ACATCGCGC.....CAACGCGCTATTTCGTC 4062
      |||
1784 ThrGlnAlaSerThrAspLysLysGlnAlaValAspGlnAlaLeuGlnAl 1800
4063 CAAAAGCGGATTACCGAT 4081
      |||
1800 aAlaGlnSerIleThrAsp 1806

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAU37403
seq_documentation_block:
ID   AAU37403 standard; Protein; 6281 AA.
XX
AC   AAU37403;
XX
DT   14-FEB-2002 (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1573.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KM   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   W0200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX

```









3788 ThrValGluGlnThrLysGlnAsnAlaGlnAsnValAsnThrAlaMetSer 3804  
4107 GGGCTTGCATTCAACCGCTACCGCGGGGCAATT..... 4140  
:::| | | | |  
3804 rAsnLeu.....LysGlnGlyIleAlaAsnLysAspThrV 3816  
4141 ..AAGCGAGATTATTCATTCAACCGCGCAACACATTTCATCAGCGCT 4188  
| | | | |:::| | | | |  
3816 allYsAlaSerGlnAsnTyrHisAspAlaAspAlaAspLysGlnThr... 3831  
4189 TATTTGAGCTGTCCTATACCGATGCC.....GCTTCGGCGCAAGT 4229  
3832 .....AlaTyrThrAsnAlaValSerGlnAlaGluGlyIleH 3844  
4230 CCGAAGCGCGTCAATACCGCGTA.....TTGGCGAGGATT 4267  
:::| | | | |:::| | | | |  
3844 eAsnGlnThrThrAsnProThrLeuAsnProAspGluIleThrAlaAlaL 3861  
4268 TGGGCAAAACCCGAGTGGGGAATGGGGGTAAACCGCAATCAAA 4314  
:::| | | | |:::| | | | |  
3861 euThrGlnValThrAspAlaLysAsnGlyLeuAsnGlyGlnAlaLys 3876

